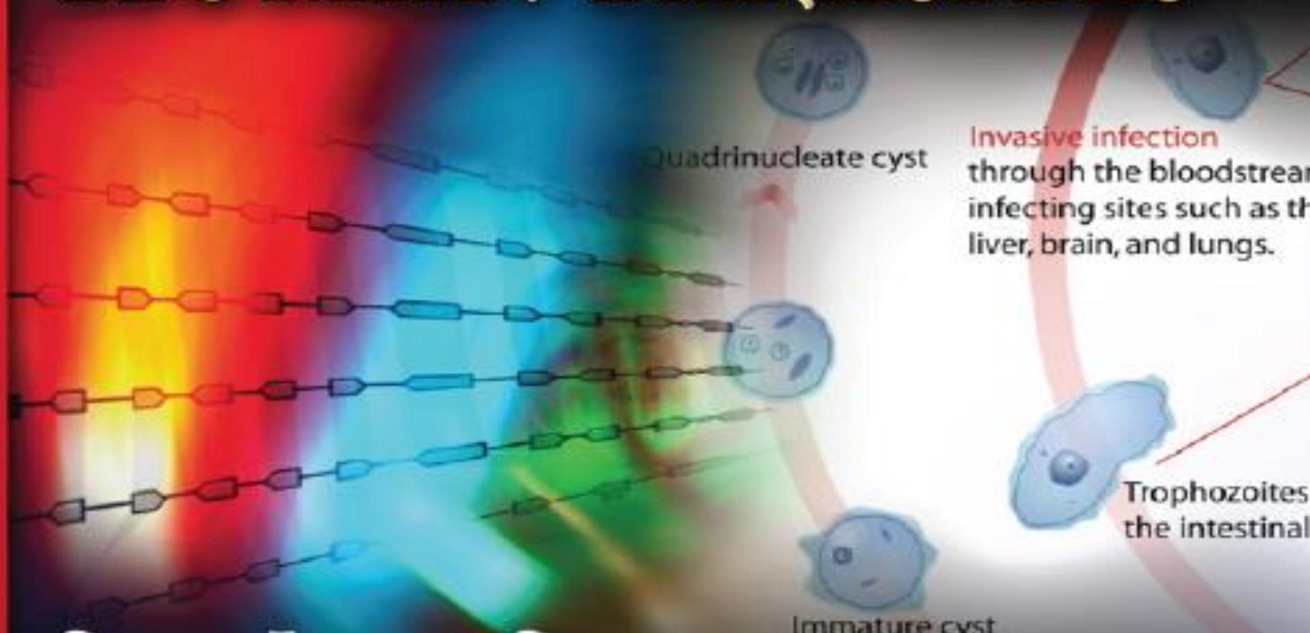
Nova Biomedical

Genetics - Research and Issues

Comparative Genomics in Neglected Human Parasites



César López-Camarillo
Laurence A. Marchat
Editors

NOVA

For the exclusive use of R. Manning-Cela

Copyright © 2013 by Nova Science Publishers, Inc.

All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means: electronic, electrostatic, magnetic, tape, mechanical photocopying, recording or otherwise without the written permission of the Publisher.

For permission to use material from this book please contact us:
Telephone 631-231-7269; Fax 631-231-8175
Web Site: http://www.novapublishers.com

NOTICE TO THE READER

The Publisher has taken reasonable care in the preparation of this book, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained in this book. The Publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance upon, this material. Any parts of this book based on government reports are so indicated and copyright is claimed for those parts to the extent applicable to compilations of such works.

Independent verification should be sought for any data, advice or recommendations contained in this book. In addition, no responsibility is assumed by the publisher for any injury and/or damage to persons or property arising from any methods, products, instructions, ideas or otherwise contained in this publication.

This publication is designed to provide accurate and authoritative information with regard to the subject matter covered herein. It is sold with the clear understanding that the Publisher is not engaged in rendering legal or any other professional services. If legal or any other expert assistance is required, the services of a competent person should be sought. FROM A DECLARATION OF PARTICIPANTS JOINTLY ADOPTED BY A COMMITTEE OF THE AMERICAN BAR ASSOCIATION AND A COMMITTEE OF PUBLISHERS.

Additional color graphics may be available in the e-book version of this book.

Library of Congress Cataloging-in-Publication Data

ISBN: 978-1-62417-881-8 (eBook)
Library of Congress Control Number: 2012956337

Published by Nova Science Publishers, Inc. † New York

CONTENTS

Preface		**vii**
Chapter 1	Protein Families in *Plasmodium* Parasites: A Network Perspective *Timothy G. Lilburn, Yufeng Wang, Hong Cai and Jianying Gu*	**1**
Chapter 2	*Trypanosoma cruzi* Genome: Organization, Dynamics, Function and Promise *C. Lugo-Caballero, D. Sánchez-Cruz, L.A. Hernández-Osorio, G. Noris-Sarabia, M. Rubio-Ortiz, C. Marquez-Dueñas, S. Martinez-Calvillo and R. Manning-Cela*	**19**
Chapter 3	Harvesting the Genome of *Trypanosoma cruzi* for the Development of New Treatments and Vaccines *Liliana Villanueva-Lizama and Eric Dumonteil*	**39**
Chapter 4	Comparative Genomics of *Leishmania* Parasites *Norma E. Padilla-Mejía, Claudia M. Gómez-Hurtado, Inti I. Sánchez-Santamaría, Luis E. Florencio-Martínez, Rebeca G. Manning-Cela and Santiago Martínez-Calvillo*	**59**
Chapter 5	Genome Architecture and Dynamics in the Protozoan Parasite *Leishmania* *Yvon Sterkers, Laurence Lachaud, Olivia Touzet, Nathalie Bourgeois, Lucien Crobu, Michel Pagès and Patrick Bastien*	**81**
Chapter 6	Insight into Genomics of *Trichomonas vaginalis* *Satendra Singh, Atul Kumar Singh, Budhayash Gautam, Prashant A. Jain and Gulshan Wadhwa*	**101**
Chapter 7	Comparative Genomics of Polyamines Pathway in Amitochondriate Excavata *Trichomonas* *María Elizbeth Alvarez Sánchez, Laura Itzel Quintas Granados, César López Camarillo and Bertha Isabel Carvajal Gamez*	**115**

For the exclusive use of R. Manning-Cela

Chapter 8	Comparative Genomics for the Identification of RecQ DNA Helicases in Protozoan Parasites *Ma. del Socorro Charcas-López, Mavil López-Casamichana, César López-Camarillo and Laurence A. Marchat*	**135**
Chapter 9	Extreme Genome Reduction in Microbial Parasites *Claudio H. Slamovits*	**157**
Index		**175**

For the exclusive use of R. Manning-Cela

In: Comparative Genomics in Neglected Human Parasites ISBN: 978-1-62417-874-0
Editors: César López-Camarillo and L. A. Marchat © 2013 Nova Science Publishers, Inc.

Chapter 2

TRYPANOSOMA CRUZI GENOME: ORGANIZATION, DYNAMICS, FUNCTION AND PROMISE

C. Lugo-Caballero[a], D. Sánchez-Cruz[a], L.A. Hernández-Osorio[a,b], G. Noris-Sarabia[a], M. Rubio-Ortíz[a], C. Marquez-Dueñas[a], S. Martínez-Calvillo[c] and R. Manning-Cela[a]*

[a]Departamento de Biomedicina Molecular, Centro de Investigación y de Estudios Avanzados del IPN, Col. San Pedro, Zacatenco, México, D.F., México

[b]Facultad de Ciencias Químicas, Universidad Autónoma Benito Juárez de Oaxaca. Oaxaca, México, Colonia Ex-hacienda Cinco Señores, Oaxaca, México

[c]Unidad de Biomedicina, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Col. Los Reyes Iztacala, Tlalnepantla, Edo. de México, México

ABSTRACT

The complete genome sequence of *Trypanosoma cruzi*, the causative agent of American trypanosomiasis, was published seven years ago. The selected strain, CL Brener, is a natural hybrid of the *T. cruzi* II and *T. cruzi* III lineages. The sequence analysis revealed that this parasite contains a diploid genome of between 106.4 and 110.7 Mb that is organized into 41 chromosome pairs and 22,570 predicted protein-coding genes, of which 12,570 represent allelic pairs. Similar to other trypanosomatids, *T. cruzi* is characterized by its unique mechanisms of gene expression, such as constitutive polycistronic transcription and *trans*-splicing. The parasite genome is organized into large polycistronic clusters of unrelated genes that are arranged sequentially on the same DNA strand, which emphasizes the importance of posttranscriptional regulation. *T. cruzi* contains a large number of repetitive sequences (≥ 50% of the genome) and includes large gene families of surface proteins (e.g., TS, mucins, gp63s and MASP), retrotransposons and subtelomeric repeats. Few promoter sequences have been identified, and general transcription factors are almost unrepresented. Therefore, little is known concerning transcription initiation and regulation. The parasite lacks classical RNA polymerase II promoter sequences. However, it has been shown that transcription

* Corresponding author: E-mail: rmanning@cinvestav.mx; Phone: (+52) (55) 57473322; Fax: (+52) (55) 57473938.

For the exclusive use of R. Manning-Cela

initiation and termination regions are epigenetically marked by histone variants or modified histones, similarly to other eukaryotes. The chromatin exhibits differential condensation through the parasite life-cycle stages, and despite the presence of nucleosomes, chromatin never folds into 30-nm fibers. In this chapter, we intend to detail the general knowledge of the *T. cruzi* genome and discuss the most recent discoveries regarding the regulation of its genome expression and the impact that these findings have on parasite biology and disease pathogenesis.

1. INTRODUCTION

Trypanosoma cruzi is the flagellated protozoan causative agent of Chagas' disease or American trypanosomiasis, which affects approximately 18 million people, and more than 100 million people are at risk of infection in endemic countries. This disease is endemic in Latin America and is transmitted to vertebrate hosts by insect vectors [1] . *T. cruzi* is an intracellular protozoan that undergoes a complex biphasic life cycle, which includes the following four distinct developmental stages: two in the reduviid beetle vector (i.e., epimastigotes and metacyclic trypomastigotes) and two in the mammalian host (i.e., amastigotes and blood trypomastigotes). In the beetle, the flagellated epimastigote proliferates in the midgut before differentiating into the nondividing but infectious metacyclic trypomastigote that is found in the vector's hindgut. The parasite infects host cells after its introduction into the mammalian blood, differentiates into an amastigote, and initiates replication in the cytosol of the infected cell. Ultimately, the amastigotes develop into nondividing bloodstream trypomastigotes, which can either initiate another round of infection or be ingested by the reduviid vector during a blood meal. The life cycle is completed upon the development of epimastigotes from bloodstream trypomastigotes (figure 1) [2; 3].

Chagas' disease presents variable clinical manifestations that include an acute phase followed by a chronic phase that is separated into two periods. In the first period, the infected individuals are relatively asymptomatic during 10-30 years, after which approximately 30% of the patients develop the clinical period, which is characterized by the presence of chagasic cardiomyopathy or gastrointestinal disorders, such as esophagus and colon dilations, although a variety of other organs may be affected (figure 1). This broad spectrum of clinical symptoms could be caused by parasitic and host genetic factors [1; 3]. Currently, no drugs or vaccines are available to cure Chagas' disease. Therefore, the main strategy for disease control is the prevention of transmission by the insect vectors and blood transfusions.

T. cruzi is a complex taxon with a remarkable genetic heterogeneity [4; 5]. Natural populations of the parasite experience clonal evolution with rare events of genetic recombination [6], although hybrid lineages have been identified in natural populations [7-10]. Using genetic and biochemical markers, *T. cruzi* strains were divided into *T. cruzi* I (TcI) (sylvatic cycle) and *T. cruzi* II (TcII) (domestic cycle) major lineages [11-13]. Further analysis led to the subdivision of the TcII lineage into the following five subgroups: Tc IIa, IIb, IIc, IId, and IIe [14]. Thereafter, phylogenetic analyses of the *T. cruzi* strains have become more complex when additional data have indicated the existence of three major groups, and hybrid strains [7; 15; 16]. Therefore, in 2009, an international consensus recognized the existence of six major strains or discrete typing units (DTUs) termed TcI - TcVI that present differential but not entirely exclusive geographical, ecological, transmission

For the exclusive use of R. Manning-Cela

cycle and disease associations [17] (Table 1). These lineages demonstrate a variety of phenotypes and present differences in their preferred host and vectors, invasion mechanisms, target cell effects, levels of parasitemia and ability to cause widely varied clinical manifestations with the consequent significant differences in their pathogenic potential, although the implications of those differences remain unclear [18]. For example, TcI isolates are less pathogenic to humans, display a lower acute infection profile and progression, which results in a more extensive chronic profile, and invade and cause pathology in different organs.

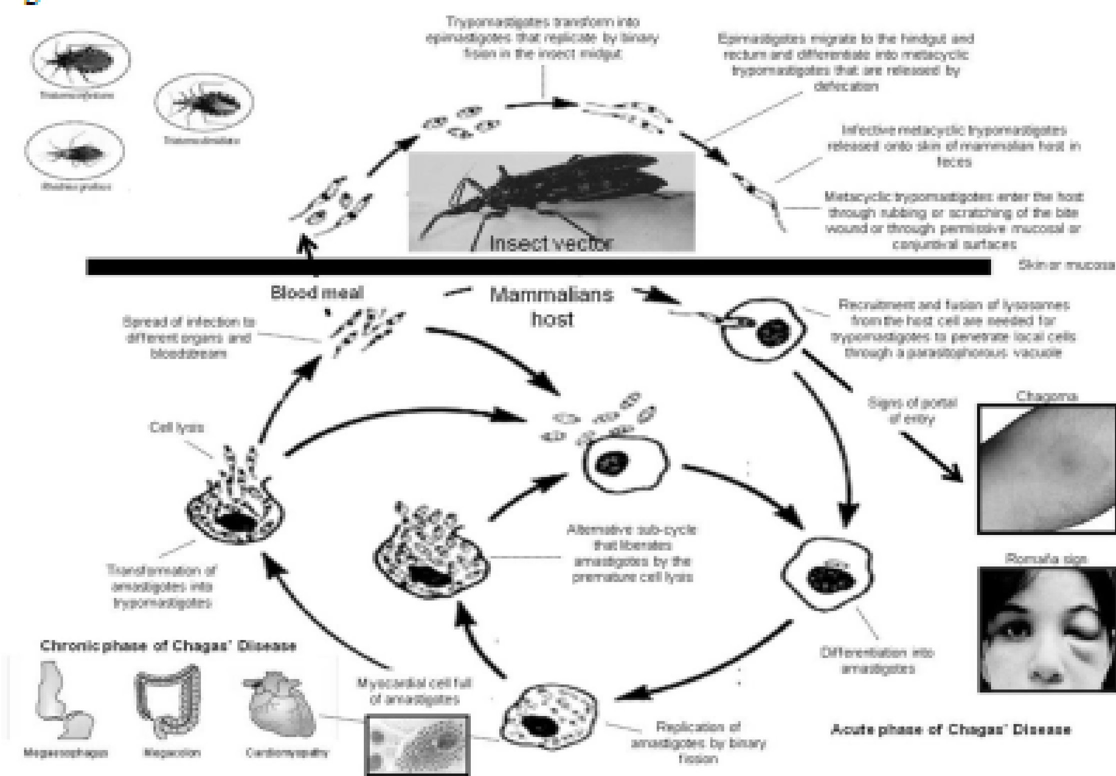


Figure 1. *T. cruzi* life cycle.

Table 1. *T. cruzi* strain classifications.

Former classification[a]	Current classification[b]	Representative strains
T. cruzi I/ DTU I	*T. cruzi* I	Sylvio X-10, Dm28c
T. cruzi II / DTU IIb	*T. cruzi* II	Esmeraldo, Y
T. cruzi III / DTU IIc	*T. cruzi* III	CM17
DTU IIa	*T. cruzi* IV	CanIII
DTU IId	*T. cruzi* V[c]	SO3
DTU IIe	*T. cruzi* VI[c]	CL Brener

DTU, discreet typing unit from [a]Momem (1999).

T. cruzi I and II classifications from Brisse *et al* (2000).

DTU I, IIa-e classification from de Freitas *et al* (2006).

T. cruzi I, II and III classification from [b]Zingales *et al* (2009).

[c]Hybrid strains.

For the exclusive use of R. Manning-Cela

Because *T. crusi* isolate differences are attributable to genetic factors, it is important to determine the genes or gene networks that could confer these different phenotypes and establish the genetic basis of the variety of clinical manifestations [19]. The availability of the *T. crusi* genome sequence provides an important platform for large-scale investigations and will allow the possibility to perform massive parallel sequencing of additional *T. crusi* strains at a fraction of the time and cost of sequencing the reference genome. Therefore, this approach can be useful to elucidate the different pathologies found in Chagas' disease and their possible association with different *T. crusi* lineages using comparative sequencing studies by pathogenomics. In addition, such comparative sequencing analyses may also contribute to developing new linage-specific diagnostic tests, performing studies of strain-specific susceptibility or resistance to drugs, executing epidemiological and evolution studies, developing new drugs, and providing further insight into the basic biology and genetics of *T. crusi*.

2. Trypanosoma Cruzi Genome

2.1. Nuclear CL Brener Genome

The diploid genome of the strain CL Brener of *T. crusi* has been estimated to be 87 Mb using pulse field gel electrophoresis assays (PFGE) and between 106.4 and 110.7 Mb according to the genome project (figure 2) [20-22]. Sequence and scaffolding genome data determined the presence of 41 pairs of chromosomes [23], although other approaches indicate that this number in *T. crusi* can be strain specific [20; 24; 25]. A high degree of synteny is characteristic of the trypanosomatid genomes, which may show synteny as high as 94% in clusters of orthologous genes (COGS) and decreasing toward the subtelomeric and telomeric regions, which are hotspots for recombination events [21].

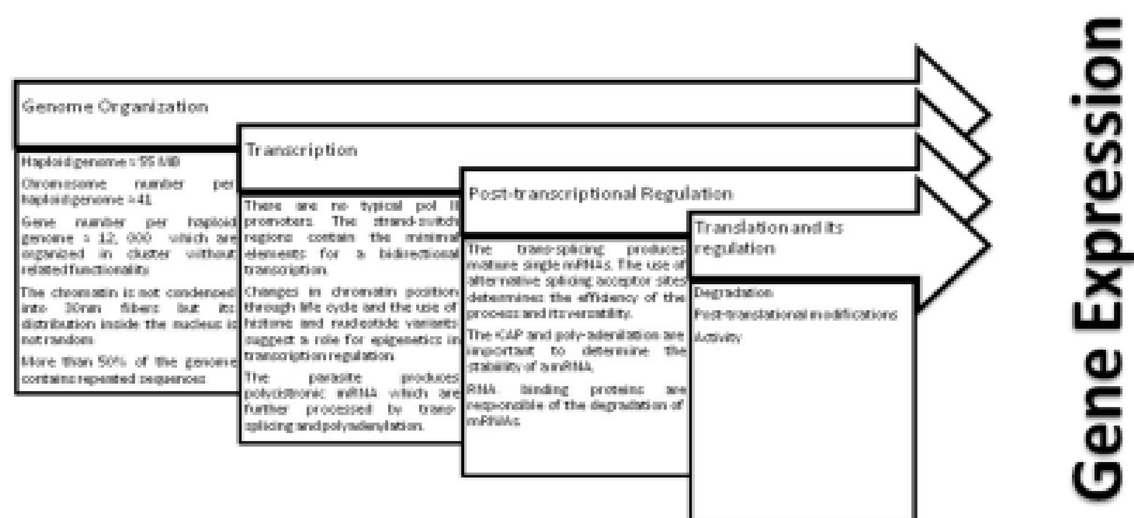


Figure 2. The genomic characteristics and gene expression of *T. crusi*.

For the exclusive use of R. Manning-Cela

The *T. cruzi* genome project estimated that the *T. cruzi* genome has 22,570 open reading frames (ORFs), and a putative function for only 50.8% of the ORFs has been predicted based on homology comparisons. Protein-coding genes are arranged in clusters that comprise hundreds of genes that may have unrelated functions, and more than 50% of the genome consists of repeated sequences that include transposable elements (TEs), surface proteins and subtelomeric repeats [21]. The repetitive region data are not accurate because of the hybrid nature of the reference strain CL Brener (TcVI results from a TcII/TcIII hybridization), which requires sequencing and analysis of the Esmeraldo TcII strain to correctly complete the work [21]. Nevertheless, genomic and post-genomic study data are being assembled in the TcruziDB, which is a freely available online website that is supported with experimental data from the research community [26; 27].

TEs are fragments of DNA that can switch positions in the genome and play an important role in the plasticity of the genome [28-30]. They are classified into the following two major groups based on the mechanism used for their transposition: 1) class I TEs or retroelements transpose via reverse transcription of an RNA intermediate and are divided into long terminal repeats (LTRs), retrotransposons and non-LTR retroposons, and 2) class II TEs move strictly through a DNA intermediate [31]. *T. cruzi* has several copies of class I TEs, including retroposons of the groups CRE and *Ingi*. The CRE group includes the retroposon CZAR that contains two different ORFs, which code for the reverse transcriptase and endonuclease that are required for its mobilization. From the *Ingi* group, the parasite genome contains the LITc, NARTc and DIRE retroposons, which are dispersed in the genome with relative site specificity [32; 33]. *T. cruzi* also has retrotransposons that are members of the DIRS/Ngaro group [34]. Finally, the genome also contains the retrotransposons, such as VIPER and the small derivative SIRE [33; 35]. The presence of abundant TEs in the genome of *T. cruzi* could be explained by the absence of a full RNAi machinery to control its duplication. A clear function for these elements has not been established, but it has been suggested that they could be important for regulating the expression of the genome [21; 31; 36]. Retroposons are an example of a gain of function because they have been shown to have DNA repair functions *in vivo*, which suggests that they could have a role in the parasite DNA repair process [37]. Multiple SIRE copies have been found in the polypyrimidine tracts of mRNAs from ribosomal proteins and in polyadenylation regions from upstream mRNAs. Moreover, VIPER retrotransposons are located in strand switch regions (SSR) of several chromosomes, which are regions in which the direction of the transcription changes. These data suggest that TEs play an important role in the plasticity of the genome [28-30].

The large gene families of surface proteins, such as mucins, trans-sialidases, mucin-associated surface proteins (MASP) and pseudogenes, are involved in the variation of the sequence repertoire through recombination events [21; 38]. Because these proteins are important for the protection and infection processes of the parasite, it has been hypothesized that their genes are repeated to assure a high level of protein expression [21; 39]. Finally, the subtelomeric repeats, as previously mentioned, are important for genome recombination [21].

For the exclusive use of R. Manning-Cela

2.2. Mitochondrial CL Brener Genome

The mitochondrial genome of kinetoplastids, known as kinetoplast DNA (kDNA), is localized near the insertion of the flagellum [40]. In *T. cruzi,* kDNA is a highly structured disk-shaped network of thousands of concatenated minicircles that are 0.5-10 kb in size and dozens of concatenated maxicircles that are 20-40 kb in size. The minicircle sequences are present exclusively in kinetoplastids, whereas maxicircles are homologs of mtDNA molecules that are found in other eukaryotes [41]. In 2006, Westenberger *et al* reported the complete sequences of maxicircles of the TcII and TcIII lineages from the Esmeraldo and CL Brener strains, respectively [42]. As in other trypanosomatid mitochondrial genes, sequence analyses indicated that there were frameshift errors in most of the genes of *T. cruzi* maxicircles. It was later demonstrated that such errors were corrected at the RNA level by a complex U-insertion/deletion process known as RNA editing [43]. Important elements of this repair process include guide RNAs (gRNAs), which are encoded mainly by minicircles and in a few cases by maxicircles. The gRNAs hybridize to the 3' end of the target message and undertake direct U insertion and deletion by the editosome machinery [44]. The complete sequences of the 25 kb *T. cruzi* maxicircles revealed 18 tightly clustered mitochondrial protein-coding genes and two rRNA genes that were syntenic with previously sequenced maxicircles of *T. brucei* and *Leishmania tarentolae*. Fifteen of the 18 protein-coding genes have been edited, and strain-specific repetitive regions and a variable region unique for each strain have been identified outside of the coding region [42].

2.3. Sylvio X10 Draft Sequence

As mentioned previously, the hybrid nature of the CL Brener genome provides data from two genomes, i.e., "Esmeraldo-like" and "non-Esmeraldo" contigs, which facilitated differentiating the TcII and TcIII groups, respectively (Table 1). Thereafter, the draft sequence of Sylvio X10, which is a representative strain of TcI, was published [45].

The Sylvio X10 genome is approximately 44 Mb, which is smaller than the CL Brener genome and characteristic of TcI strains [45; 46]. A comparison of the CL Brener and Sylvio X10 genomes indicated that the architecture of the two genomes is similar, with highly conserved syntenic regions that correspond to the gene-dense core of the coding regions that are organized into long polycistronic clusters. Similarly, with the CL Brener genome, the presence of repetitive sequences indicated that the Sylvio X10 genome was represented as fragmented contigs. The technical difficulties associated with the assembly of repetitive sequences indicated that only approximately 49% of the generated Sylvio X10 sequence data was incorporated into contigs, which resulted in 710,109 reads that were not included in the assembly. Therefore, the draft genome of Sylvio X10 was assembled into 7,092 contigs, which is slightly less than the number of contigs reported for the CL Brener draft genome. The contig alignment of both CL Brener haplotypes demonstrated that the mean nucleotide identity was greater between Sylvio X10 and non-Esmeraldo (98.2%) than between Sylvio X10 and Esmeraldo (97.5%). This finding is concordant with previous phylogenetic analyses indicating that sequences from TcI strains are more closely related to TcIII strains (represented by the non-Esmeraldo CL Brener haplotype) than with TcII strains (represented by the Esmeraldo-like haplotype) [47; 48]. The diploid Sylvio X10 genome was homozygous

For the exclusive use of R. Manning-Cela

(< 0.08% heterozygosity); this is in contrast to the hybrid CL Brener genome in which the amount of heterozygosity in the core genome was estimated to be 5.5% [21]. In addition, the analysis of the core gene content of both sequenced strains revealed six open reading frames that were missing in the Sylvio X10 genome. Several multicopy gene families, including *DGF*, *mucin*, *MASP* and *GP63* contained substantially fewer genes in Sylvio X10 than in CL Brener. The 5.9 Mb size difference between the Sylvio X10 and CL Brener genomes reflected the expansion of these gene families, although the relation of this size difference to strain differences in host preference and different clinical manifestations remains to be determined.

2.4. Mitochondrial Sylvio X10 Genome

After the publication of a maxicircle sequence from Sylvio X10, comparative analyses of the mitochondrial genomes of TcI, TcII, and TcIII were generated [48]. Consistent with the nuclear genomic analysis, the phylogenetic analysis of the maxicircle coding regions indicated a close evolutionary relationship between TcI and TcIII. Based on this analysis, a model was proposed in which an ancestral strain belonging to *T. cruzi* I provided the maxicircle for the progeny of a TcI-TcII hybridization event that resulted in the generation of TcIII and TcIV strains. The subsequent backcross hybridization between TcII and TcIII strains resulted in the TcV and TcVI strains, such as CL Brener, that carry the maxicircle from their TcIII ancestor.

3. Regulation of Gene Expression

3.1. Transcription

As in other trypanosomes, the *T. cruzi* genome is transcribed mostly by RNA polymerase II into long polycistronic primary transcripts that are composed of ten to hundreds of unrelated genes that are processed through coupled *trans*-splicing and polyadenylation to obtain mature individual mRNAs (figures 2 and 3) [49]. The rapid primary transcript processing causes the loss of the 5' start site, which makes finding promoters in this parasite difficult. Moreover, *T. cruzi* does not appear to contain typical RNA polymerase II promoters in its genome. Therefore, little is understood regarding how transcription is initiated and regulated in *T. cruzi*. *Leishmania*, a related trypanosomatid, has SSRs that contain the minimal elements necessary to initiate transcription in both directions. However, except for the presence of a C-rich tract, this region lacks recognizable eukaryotic promoter elements [50-52]. Because trypanosomatids share many gene organization and regulation characteristics, it would not be surprising if similar SSR functionality existed in *T. cruzi*. Notably, a recent genome-wide study demonstrated the enrichment of histone modifications that are associated with transcription initiation in other organisms at SSR in *T. cruzi* [53].

In *T. cruzi*, Pol II promoters are the only promoters that have been reported to facilitate SL RNA [54] and LTc transcript expression [55].

For the exclusive use of R. Manning-Cela

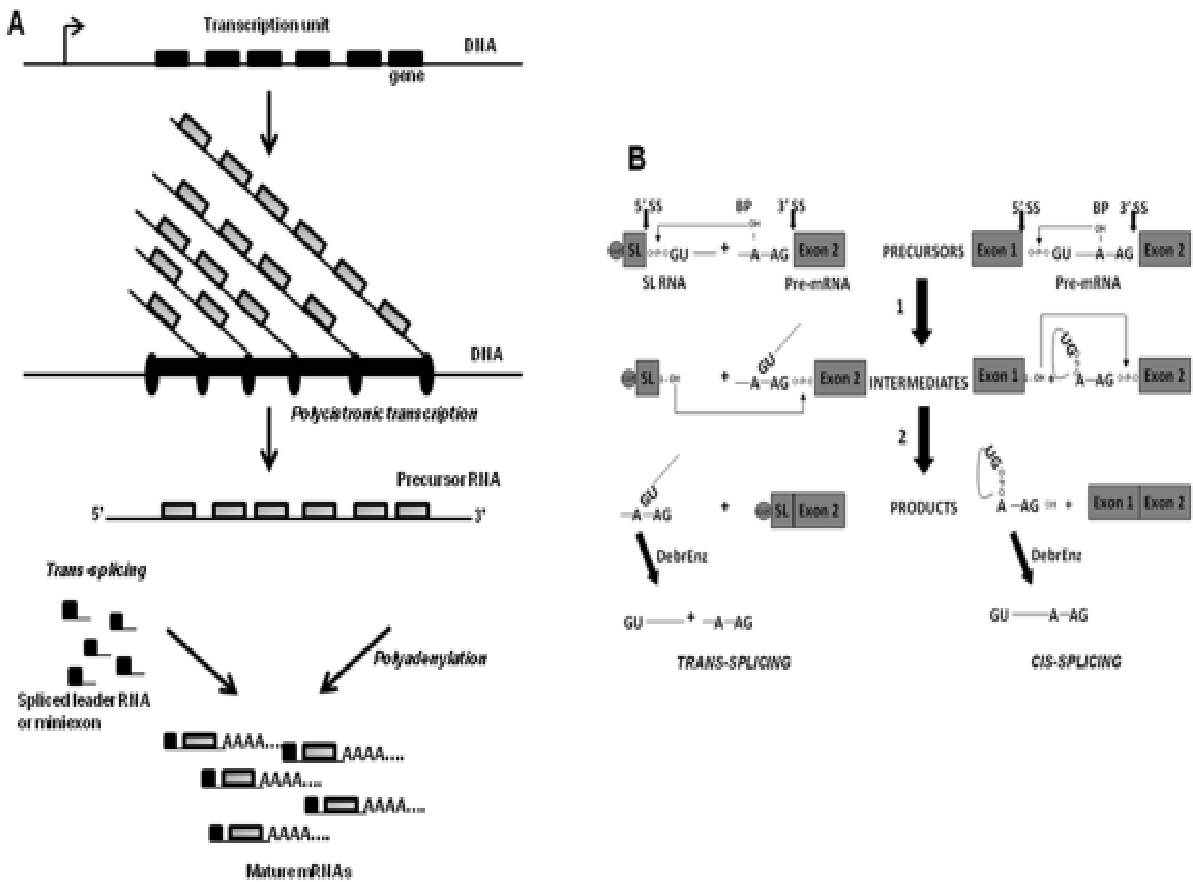


Figure 3. Polycistronic transcription and mRNA processing. (A) The *T. cruzi* genome is arranged into polycistronic clusters of several genes that are transcribed by a single promoter-like region to produce a primary transcript that is processed by *trans*-splicing and polyadenylation to generate mature mRNAs. (B) This schematic represents pre-mRNA *cis*-splicing and *trans*-splicing pathways. The exons are shown inside boxes. The introns are indicated by solid lines. The circle represents the cap 4 structure of the SL RNA. GU is a conserved dinucleotide at the 5' splice site (5' SS). AG, conserved dinucleotide at the 3' splice site (3' SS). BP, branch point. DebrEnz, debraching enzyme.

For the exclusive use of R. Manning-Cela

3.2. Epigenetic Control

Despite the apparent absence of typical promoters in *T. cruzi*, several lines of evidence suggest that epigenetic mechanisms play an important role in gene expression regulation in this parasite. For example, many posttranslational histone modifications, histone variants and chromatin-modifying enzymes have been identified in *T. cruzi* and other trypanosomatids [24; 53; 56-61].

T. cruzi has several copies of the genes that encode histones H1, H2A, H2B, H3 and H4. However, histones in this parasite are extremely divergent from those found in other organisms. Nevertheless, as in other organisms, nucleosomes constitute the basic structural unit of chromatins in trypanosomatids. Histone H1 lacks the globular domain [62], thus contributing to the limited compaction of the trypanosome chromatin with no further packing into 30 nm-fibers and metaphase chromosomes [63]. Altered chromatin condensation is observed at different developmental stages of *T. cruzi* because higher compaction is observed in the non-replicative trypomastigote stage compared with the epimastigote stage. Moreover, variation in total mRNA levels is observed during development, as well as an increased amount of heterochromatin, which suggests a possible epigenetic contribution to transcription [57].

In this parasite, histones H4 and H2A are mainly acetylated, whereas histones H3 and H2B are preferentially methylated [56; 58-61]. Histone H4, which shares 96% homology between *T. cruzi* and *T. brucei*, is abundantly acetylated at Lys4 (K4) and to a lesser extent at K10 and K14. Histone H4 is additionally acetylated at K2 and K5 in *T. brucei* and at K57 in *T. cruzi*. K4 acetylation occurs preferentially in the proliferative stages and accumulates when chromatin is packaged, and K10 and K14 acetylation have a particular distribution confined between packed and open chromatin [64]. Additionally, a recent report revealed that acetylated (H3-K9/K14 and H4-K5/K8/K12/K16) and methylated (H3-K4) histones are enriched at divergent SSR in the *T. cruzi* genome [53].

Similarly, a ChIP-seq analysis in *T. brucei* demonstrated that histone variants H2AZ and H2BV are enriched at probable Pol II transcription start sites and that histone variants H3V and H4V are enriched in regions in which transcription appears to terminate [65]. These findings suggest that posttranslational histone modifications and histone variants play an important role in transcription initiation and termination in trypanosomatids [66; 67].

Recent studies in *T. cruzi* indicate that base J (β-D-glucosyl-hydroxymethyluracil) is an epigenetic factor that regulates Pol II transcription initiation via DNA modification, chromatin remodeling and increased occupancy of Pol II [68; 69]. The hydroxylation and glucosylation of the thymine base of DNA produces the modified base J, in which down-regulation has been recently associated to nucleosome loss, increased H4 and H3 histone acetylation and recruitment of Pol II, which causes an improved transcription initiation rate. In JBP1 null and JBP2 null parasites, enzymes that regulate base J synthesis, exhibited a defect in the expression of genes that are involved in *T. cruzi* virulence, which causes defective behavior in parasite invasion and departure from host cells and the loss of base J at sites of transcription initiation [69]. Despite all of the evidence that has been accumulated to date, little is known concerning transcription initiation in *T. cruzi*. Further work is required to comprehend the atypical mechanisms of gene expression in trypanosomatids.

For the exclusive use of R. Manning-Cela

3.3. Posttranscriptional Gene Regulation

A low proportion of the other trypanosomatid genomes is stage regulated [70-72]. In contrast, *T. cruzi* displays stage-regulated control of mRNA abundance for more than 50% of its genes [73]. Besides, the *T. cruzi* genome is arranged into polycistronic clusters of several genes that are transcribed by a single promoter-like region and presents epigenetic regulation. Therefore, it has been hypothesized that its gene expression is mainly regulated posttranscriptionally, i.e., at the level of *trans*-splicing efficiency, mRNA stability and decay, translation efficiency, and posttranslational modifications (figure 2) [74; 75].

3.3.1. Trans-Splicing

Trans-splicing, a mechanism described 30 years ago in trypanosomatids, is the processing of polycistronic RNA to mature monocistronic mRNA through two consecutive transesterification reactions that utilize two RNA molecules, the spliced leader (SL) RNA and the pre-mRNA (figure 3). An AG dinucleotide at the 3' splice site and an upstream pyrimidine-rich region are the most conserved sequences required for this process [76-78]. In the first step of the transesterification reaction, there is a cleavage at the 5' splice donor site on the SL RNA that produces a Y-branched intermediary structure, instead of a lariat typical of *cis*-splicing. During the second step of the reaction, the Y-branched intermediate is removed when the SL sequence is ligated to the pre-mRNA at the 3' splice acceptor site (SAS) (figure 3) [79]. The polyadenylation of the mRNA relies on this reaction and on the downstream splicing efficiency of its adjacent mRNA [76; 79] . Moreover, the efficiency of all these splicing signals is contingent on their sequence composition and length because the mRNA is poorly expressed when the degradation of a given untranslated region (UTR) is faster than the splicing reaction [78; 80]. Another important factor for *trans*-splicing efficiency is the length and position of the polypyrimidine tract relative to the 3' splice acceptor site [81; 82] . Regularly, the first canonical SAS downstream of the polypyrimidine tract is used by *trans*-splicing factors following a scanning model [77], although less efficient non-canonical SASs have been reported [83].

Alternative *trans*-splicing has been demonstrated in trypanosomes [83-87]. As an example, in *T. cruzi, LYT1* gene expression can produce alternative transcripts [83] that encode dually targeted proteins involved in parasite infection and stage differentiation phenotypes that demonstrate differential dual localization in the plasma membrane and the base of the parasite flagellum [83; 84].

3.3.2. Stability and Decay of mRNAs

The stability and decay of the mRNA in the cytoplasm rely on signals present on 3'- and 5'-UTRs and several RNA-binding proteins that are involved in its recognition and degradation pathways. For instance, the presence of a specific 3'-UTR sequence stabilizes the transcript of the small mucin gene (*SMUG*) in *T. cruzi* [88-90]. Few RNA-binding proteins involved in stability have been characterized, such as the poly-A binding protein (PABP1) that interacts with the 3'-UTR of mRNAs to confer stability and the U-rich RNA-binding protein (TcUBP1) that confers instability to the *SMUG* mRNA [91; 92].

For the exclusive use of R. Manning-Cela

3.3.3. Protein Translation and Turnover

Another important level of posttranscriptional regulation in this parasite occurs during protein translation and turnover. Whole proteome characterization of *T. cruzi* has provided insight into the proteins involved in development and distinct metabolic pathways [38; 93]. Fractionation or enrichment of specific organelles to obtain subcellular proteomes has been performed to detect underrepresented proteins following extensive and complex purification [94]. Therefore, subproteomic *T. cruzi* studies have been performed in reservosomes [95; 96], glycosomes [97] and GPI-anchored proteins [98]. Additionally, posttranslational modifications, such as phosphoproteome (kinome) [99] and histone acetylation and deamidation [100], have been recently reported.

4. Gene and Protein Families

T. cruzi-specific genes are contained at non-syntenic chromosome-internal and subtelomeric regions and consist mainly of members of large families of surface antigens, such as trans-sialidase (TS), mucin-associated surface proteins and mucins (TcMUC). These gene families are clustered into large arrays that are as large as 600 Kb and have been subjected to intense rearrangements through the evolution of the parasite [21; 101]. Therefore, it is possible that many of the polymorphisms observed among *T. cruzi* isolates, which are exhibited in numerous epidemiological and pathological characteristics of Chagas' disease, are in part due to the variability within these regions [19]. The expansion of these gene families, structural RNAs and retroelements is often associated with breaks in synteny. Gene divergence, acquisition and loss of genes, and rearrangements within and between syntenic regions have shaped the genomes of the trypanosomatids [102].

The TS superfamily, which is the largest gene family of *T. cruzi*, consists of 1430 genes (and 693 pseudogenes) that are located in subtelomeric and interchromosomal regions, are associated with other genes and code for surface proteins [21]. This superfamily is characterized by the conserved motif VTVxNVxLYNR, as well as other less conserved motifs, such as Asp box and FRIP (xRxP) [103]. This superfamily has been classified into the following two subfamilies: 1) TS active, which consists of 12 members that contain the active residue Try342 required for enzymatic activity; most members have a variable number of tandem repeat sequences (e.g., 12 amino acids) termed SAPA (shed acute-phase antigen) and are GPI anchored; and 2) TS-like, which comprises 725 members that have no TS activity because they lack the active residue Try342 that is replaced by His342 [21; 104]. Additionally, TSs are classified into four subfamilies or groups according to their sequence similarity and functional properties. Group I comprise active TSs, such as TCNA and SAPA (i.e., expressed in trypomastigote) and TS-epi (i.e., expressed in epimastigote). Group II comprises gp85 surface glycoproteins that are involved in the binding and invasion of host cells (e.g., TSA-1, SA85, gp90, gp82 and ASP-2). Group III is represented by the regulatory protein FL-160 that has been associated with the inhibition of classical and alternative pathways of the complement system. Finally, Group IV is composed of TsTc13, which is a protein with an unknown function but is classified within the TS superfamily because it contains the conserved motif VTVxNVxLYNR [103; 105]. Recently, this classification was extended to eight groups (I to VIII) according to their TS motif (FRIP motif, Asp box, more

For the exclusive use of R. Manning-Cela

than 10 amino acids repeated and canonical motif VTVxNVxLYNR), chromosomal location, expression profile and antigenic properties. This study analyzed 508 complete TS sequences and excluded partial sequences and pseudogenes. Groups I, II, IV, V and VI possess Asp boxes; groups I, III, IV, VII and VIII have the FRIP motif; groups I and IV have an amino acid repeat at the N-terminus; and all groups exhibited the canonical motif VTVxNVxLYNR. The chromosomal localization of the groups is as follows: 60 genes are found in subtelomeric regions (36 members of group II, 7 of group IV and 10 of group VIII). Interestingly all members of group IV (the largest one) are located in interchromosomal regions. In conclusion, genomic and functional studies have revealed that the variability of the TS family is higher than previously understood [103].

The *T. cruzi* genome project also described a new family of genes related to mucins called mucin-associated surface proteins (MASP). This family is the second largest in the genome and is composed of 1377 genes (771 are complete genes, and 433 are pseudogenes) that are regularly located downstream of TcMUC II, mainly in large chromosomes. This family is characterized by the presence of highly conserved N- and C-terminal domains that encode a signal peptide and GPI anchor, respectively, as well as a hypervariable central region (both in sequence and length) with repeating patterns. Unlike the heterogeneity of the coding region, the 5'- and 3'-UTRs are highly conserved [21].

The mucin family represents the third largest family of surface proteins in *T. cruzi*. This family is classified into the subfamilies TcMUC and TcSMUG based on a comparison of their sequences. The TcMUC subfamily is composed of 844 members that are organized in head-to-tail tandem repeats or interspersed in the genome. Their members are expressed in blood trypomastigote and amastigote forms and are subdivided in two groups, TcMUC I and TcMUC II. TcMUC I is characterized by the internal tandem repeat sequence Thr-Lys-Pro, which is flanked by N- (signal peptide) and C-terminal (GPI anchor) domains. Group TcMUC II, unlike TcMUC I, has noncentral variable regions that are rich in threonine, serine and proline. The *TcSMUG* subfamily or small mucin-like genes are composed of 19 members that are organized in tandem arrays and are relatively more homogeneous than the subfamily *TcMUC*. These genes are expressed in epimastigotes and are divided into *TcSMUG S* (small) and *TcSMUSG L* (large) subfamilies based on their mRNA size [21; 39; 88; 106; 107].

Finally, there are other large gene families that are important in the biology of *T. cruzi*, such as the gp63 surface metalloprotease family that is involved in the virulence of *T. cruzi* and consists of over 420 genes, and pseudogenes dispersed throughout the genome that can also be grouped in tandem [21]. Additionally, the dispersed gene family of protein 1 (DGF-1 family) is composed of 565 genes (of which 136 are pseudogenes) that are dispersed throughout the *T. cruzi* genome and have an unknown function [21; 108]. This family is organized into at least three groups with a differential distribution of functional domains.

Conclusion

Although it has been reported that the differences between *T. cruzi* isolates are genetically programmed, it has not been established what genes or clusters of genes confer their phenotypic differences. Therefore, the different pathologies found in Chagas disease and their possible association with distinct *T. cruzi* strains requires exploration through

For the exclusive use of R. Manning-Cela

comparative sequencing by pathogenomics. Genomic sequence analyses of multiple *T. cruzi* isolates are feasible using massive parallel sequencing technologies at a fraction of the time and cost of sequencing the reference genome. The availability of these genomes sequences will facilitate performing functional studies to understand the genetic basis of the clinical variability and pathogenic features of this parasite, identifying disease biomarkers and responder molecules to diverse environmental stimulus during the parasite life cycle and discovering potential novel drug, vaccines and diagnostic targets.

Moreover, with the genomic data available, primary biochemical literature and transcriptomic and proteomic profiles, *T. cruzi* is an ideal model organism for integral studies using diverse data sources to understand the interaction between the parasite and target cells. This approach offers a valuable framework to increase our knowledge concerning parasite biology (intra-species interactions) and host-pathogen interaction mechanisms (inter-species interactions).

Acknowledgments

This work was supported by grants 132312 and 139898 from CONACyT and grant "Proyectos multidiciplinarios" from CINVESTAV-IPN to R. Manning-Cela and grants 128461 from CONACyT and IN210712 from PAPIIT (UNAM) to S. Martínez-Calvillo. M. Rubio-Ortíz was a recipient of a masters fellowship from CONACyT. C. Lugo-Caballero, D. Sánchez-Cruz, G. Noris-Sarabia were recipients of a doctoral fellowship from CONACyT. L.A. Hernández-Osorio was a recipient of a doctoral fellowship from Universidad Autónoma Benito Juárez de Oaxaca, Oaxaca, México.

References

[1] Rassi, AJr.; Rassi A; Marin-Neto JA. Chagas disease. *Lancet*, 2010 375, 1388-1402.

[2] Burleigh, BA; Andrews NW. The mechanims of *Trypanosoma cruzi* invasion of mammalian cells. *Annual Review Microbiology*, 1995 49, 175-200.

[3] Tyler, KM; Engman, DM. The life cycle of *Trypanosoma cruzi* revisited. *International Journal For Parasitology*, 2001 31, 472-481.

[4] Tibayrenc, M. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *International Journal For Parasitology*, 1998 28, 85-104.

[5] Tibayrenc, M. Genetic subdivisions within *Trypanosoma cruzi* (Discrete Typing Units) and their relevance for molecular epidemiology and experimental evolution. *Kinetoplastid Biology And Disease*, 2003 2, 12.

[6] Tibayrenc, M; Ward, P; Moya, A; Ayala, FJ. Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure. *Proceedings Of The National Academy Of Sciences USA*, 1986 83, 115-119.

[7] de Freitas, JM; Augusto-Pinto, L; Pimenta, JR; Bastos-Rodrigues, L; Goncalves, VF; Teixeira, SM; Chiari, E; Junqueira, AC; Fernandes, O; Macedo, AM; Machado, CR;

For the exclusive use of R. Manning-Cela

Pena, SD. Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. *PLoS Pathogens*, 2006 2, e24.

[8] Gaunt, MW; Yeo, M; Frame, IA; Stothard, JR; Carrasco, HJ; Taylor, MC; Mena, SS; Veazey, P; Miles, GA; Acosta, N; de Arias, AR; Miles, MA. Mechanism of genetic exchange in American trypanosomes. *Nature*, 2003 421, 936-939.

[9] Tomazi, L; Kawashita, SY; Pereira,PM; Zingales, B; Briones, MR. Haplotype distribution of five nuclear genes based on network genealogies and Bayesian inference indicates that *Trypanosoma cruzi* hybrid strains are polyphyletic. *Genetics And Molecular Research*, 2009 8, 458-476.

[10] Westenberger, SJ; Barnabe, C; Campbell, DA; Sturm, NR. Two hybridization events define the population structure of *Trypanosoma cruzi*. *Genetics*, 2005 171, 527-543.

[11] Momen, H. Taxonomy of *Trypanosoma cruzi*: a commentary on characterization and nomenclature. *Memorias Do Instituto Oswaldo Cruz*, 1999 94 Suppl 1, 181-184.

[12] Souto, RP; Fernandes, ; Macedo, AM; Campbell, DA; Zingales, B. DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. *Molecular And Biochemical Parasitology*, 1996 83, 141-152.

[13] Zingales B; Souto, RP; Mangia, RH; Lisboa, CV; Campbell, DA; Coura, JR; Jansen, A; Fernandes, O. Molecular epidemiology of American trypanosomiasis in Brazil based on dimorphisms of rRNA and mini-exon gene sequences. *International Journal For Parasitology*, 1998 28, 105-112.

[14] Brisse, S; Dujardin, JC; Tibayrenc, M. Identification of six *Trypanosoma cruzi* lineages by sequence-characterised amplified region markers. *Molecular And Biochemical Parasitology*, 2000 111, 95-105.

[15] Augusto-Pinto, L; Teixeira, SM; Pena, SD; Machado, CR. Single-nucleotide polymorphisms of the *Trypanosoma cruzi* MSH2 gene support the existence of three phylogenetic lineages presenting differences in mismatch-repair efficiency. *Genetics*, 2003 164, 117-126.

[16] Miles, MA; Souza, A; Povoa, M; Shaw, JJ; Lainson, R; Toye, PJ. Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas' disease in Amazonian Brazil. *Nature*, 1978 272, 819-821.

[17] Zingales, B; Andrade, SG; Briones, MR; Campbell, DA; Chiari, E; Fernandes, O; Guhl, F; Lages-Silva, E; Macedo, AM; Machado, CR; Miles, MA; Romanha, AJ; Sturm, NR; Tibayrenc, M; Schijman, AG. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Memorias Do Instituto Oswaldo Cruz*, 2009 104, 1051-1054.

[18] Miles, MA; Llewellyn, MS; Lewis, MD; Yeo, M; Baleela, R; Fitzpatrick, S; Gaunt, MW; Mauricio, IL. The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. *Parasitology*, 2009 136, 1509-1528.

[19] Choi, J; El-Sayed, NM. Functional genomics of trypanosomatids. *Parasite Immunology*, 2012 34, 72-79.

[20] Cano, MI; Gruber, A; Vazquez, M; Cortes, A; Levin, MJ; Gonzalez, A; Degrave, W; Rondinelli, E; Zingales, B; Ramirez, JL. et al. Molecular karyotype of clone CL Brener chosen for the *Trypanosoma cruzi* genome project. *Molecular And Biochemical Parasitology*, 1995 71, 273-278.

For the exclusive use of R. Manning-Cela

[21] El-Sayed, NM; Myler, PJ; Bartholomeu, DC; Nilsson, D; Aggarwal, G; Tran, AN; Ghedin, E; Worthey, EA; Delcher, AL; Blandin, G. et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*, 2005 309, 409-415.

[22] R. T. Souza; F. M. Lima; R. M. Barros; D. R. Cortez; M. F. Santos; E. M. Cordero; J. C. Ruiz; S. Goldenberg; M. M. Teixeira and J. F. da Silveira. Genome size, karyotype polymorphism and chromosomal evolution in *Trypanosoma cruzi*. *PLoS One*, 2011 6, e23042.

[23] Weatherly, DB; Boehlke, C; Tarleton, RL. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics*, 2009 10, 255.

[24] Hanke, J; Sanchez, DO; Henriksson, J; Aslund, L; Pettersson, U; Frasch, AC; Hoheisel, JD. Mapping the *Trypanosoma cruzi* genome: analyses of representative cosmid libraries. *Biotechniques*, 1996 21, 686-688, 690-683.

[25] Henriksson, J; Aslund, L; Pettersson, U. Karyotype variability in *Trypanosoma cruzi*. *Parasitology Today*, 1996 12, 108-114.

[26] Aguero, F; Zheng, W; Weatherly, DB; Mendes, P; Kissinger, JC. TcruziDB: an integrated, post-genomics community resource for *Trypanosoma cruzi*. *Nucleic Acids Research*, 2006 34, D428-431.

[27] Andersson, B. The *Trypanosoma cruzi* genome; conserved core genes and extremely variable surface molecule families. *Research In Microbiology*, 2011 162, 619-625.

[28] Ben-Dov, CP; Levin, MJ; Vazquez, MP. Analysis of the highly efficient pre-mRNA processing region HX1 of *Trypanosoma cruzi*. *Molecular And Biochemical Parasitology*, 2005 140, 97-105.

[29] Lorenzi, HA; Robledo, G; Levin, MJ. The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-enconding retrotransposons. *Molecular And Biochemical Parasitology*, 2006 145, 184-194.

[30] Vazquez, M; Lorenzi, H; Schijman, AG; Ben-Dov, C; Levin, MJ. Analysis of the distribution of SIRE in the nuclear genome of *Trypanosoma cruzi*. *Gene*, 1999 239, 207-216.

[31] Bringaud, F; Ghedin, E; El-Sayed, NM; Papadopoulou, B. Role of transposable elements in trypanosomatids. *Microbes And Infection*, 2008 10, 575-581.

[32] Bringaud, F; Bartholomeu, DC; Blandin, G; Delcher, A; Baltz, T; El-Sayed, NM; Ghedin, E. The *Trypanosoma cruzi* L1Tc and NARTc non-LTR retrotransposons show relative site specificity for insertion. *Molecular Biology And Evolution*, 2006 23, 411-420.

[33] Bringaud, F; Garcia-Perez, JL; Heras, SR; Ghedin, E; El-Sayed, NM; Andersson, B; Baltz, T; Lopez, MC. Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. *Molecular And Biochemical Parasitology*, 2002 124, 73-78.

[34] Aksoy, S. Site-specific retrotransposons of the trypanosomatid protozoa. *Parasitology Today*, 1991 7, 281-285.

[35] Vazquez, M; Ben-Dov, C; Lorenzi, H; Moore, T; Schijman, A; Levin, MJ. The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons. *Proceedings Of The National Academy Of Sciences USA*, 2000 97, 2128-2133.

[36] Ullu, E; Tschudi, C; Chakraborty, T. RNA interference in protozoan parasites. *Cell Microbiology*, 2004 6, 509-519.

For the exclusive use of R. Manning-Cela

[37] Olivares, M; Lopez, MC; Garcia-Perez, JL; Briones, P; Pulgar, M; Thomas, MC. The endonuclease NL1Tc encoded by the LINE L1Tc from *Trypanosoma cruzi* protects parasites from daunorubicin DNA damage. *Biochimica Et Biophysica Acta*, 2003 1626, 25-32.

[38] Atwood, JA3rd; Weatherly, DB; Minning, TA; Bundy, B; Cavola, C; Opperdoes, FR; Orlando, R; Tarleton, RL. The *Trypanosoma cruzi* proteome. *Science*, 2005 309, 473-476.

[39] Buscaglia, CA; Campo, VA; Frasch, AC; Di Noia, JM. *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nature Reviews Microbiology*, 2006 4, 229-236.

[40] Brener, Z. Biology of Trypanosoma cruzi. *Annual Review Of Microbiology*, 1973 27, 347-382.

[41] Lukes, J; Jirku, M; Dolezel, D; Kral'ova, I; Hollar, L; Maslov, DA. Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. *Journal Of Molecular Evolution*, 1997 44, 521-527.

[42] Westenberger, SJ; Cerqueira, GC; El-Sayed, NM; Zingales, B; Campbell, DA; Sturm, NR. *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. *BMC Genomics*, 2006 7, 60.

[43] Hajduk, SL; Harris, ME; Pollard, VW. RNA editing in kinetoplastid mitochondria. *The FASEB Journal* 1993 7, 54-63.

[44] Stuart, K; Panigrahi, AK. RNA editing: complexity and complications. *Molecular Microbiology*, 2002 45, 591-596.

[45] Franzen, O; Ochaya, S; Sherwood, E; Lewis, MD; Llewellyn, MS; Miles, MA; Andersson, B. Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener. *PLoS Neglected Tropical Diseases*, 2011 5, e984.

[46] Branche, C; Ochaya, S; Aslund, L; Andersson, B. Comparative karyotyping as a tool for genome structure analysis of *Trypanosoma cruzi*. *Molecular And Biochemical Parasitology*, 2006 147, 30-38.

[47] Cerqueira, GC; Bartholomeu, DC; DaRocha, WD; Hou, L; Freitas-Silva, DM; Machado, CR; El-Sayed, NM; Teixeira, SM. Sequence diversity and evolution of multigene families in *Trypanosoma cruzi*. *Molecular And Biochemical Parasitology*, 2008 157, 65-72.

[48] Ruvalcaba-Trejo, LI; Sturm, NR. The *Trypanosoma cruzi* Sylvio X10 strain maxicircle sequence: the third musketeer. *BMC Genomics*, 2011 12, 58.

[49] Campbell, DA; Thomas, S; Sturm, NR. Transcription in kinetoplastid protozoa: why be normal? *Microbes And Infection*, 2003 5, 1231-1240.

[50] Martinez-Calvillo, S; Yan, S; Nguyen, D; Fox, M; Stuart, K; Myler, PJ. Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Molecular Cell*, 2003 11, 1291-1299.

[51] Schimanski, B; Nguyen, TN; Gunzl, A. Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in *Trypanosoma brucei*. *Molecular And Cellular Biology*, 2005 25, 7303-7313.

[52] Worthey, EA; Martinez-Calvillo, S; Schnaufer, A; Aggarwal, G; Cawthra, J; Fazelinia, G; Fong, C; Fu, G; Hassebrock, M; Hixson, G. et al. *Leishmania major* chromosome 3

For the exclusive use of R. Manning-Cela

contains two long convergent polycistronic gene clusters separated by a tRNA gene. *Nucleic Acids Research*, 2003 31, 4201-4210.

[53] Respuela, P; Ferella, M; Rada-Iglesias, A; Aslund, L. Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *Trypanosoma cruzi*. *The Journal Of Biological Chemistry*, 2008 283, 15884-15892.

[54] Nunes, LR; Carvalho, MR; Shakarian, AM; Buck, GA. The transcription promoter of the spliced leader gene from *Trypanosoma cruzi*. *Gene*, 1997 188, 157-168.

[55] Heras, SR; Lopez, MC; Olivares, M; Thomas, MC. The L1Tc non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts. *Nucleic Acids Research*, 2007 35, 2199-2214.

[56] da Cunha, JP; Nakayasu, ES; de Almeida, IC; Schenkman, S. Post-translational modifications of *Trypanosoma cruzi* histone H4. *Molecular And Biochemical Parasitology*, 2006 150, 268-277.

[57] Elias, MC; Marques-Porto, R; Freymuller, E; Schenkman, S. Transcription rate modulation through the *Trypanosoma cruzi* life cycle occurs in parallel with changes in nuclear organisation. *Molecular And Biochemical Parasitology*, 2001 112, 79-90.

[58] Janzen, CJ; Fernandez, JP; Deng, H; Diaz, R; Hake, SB; Cross, GA. Unusual histone modifications in *Trypanosoma brucei*. *FEBS Letters*, 2006 580, 2306-2310.

[59] Janzen, CJ; Hake, SB; Lowell, JE; Cross, GA. Selective di- or trimethylation of histone H3 lysine 76 by two DOT1 homologs is important for cell cycle regulation in *Trypanosoma brucei*. *Molecular Cell*, 2006 23, 497-507.

[60] Mandava, V; Fernandez, JP; Deng, H; Janzen, CJ; Hake, SB; Cross, GA. Histone modifications in *Trypanosoma brucei*. *Molecular And Biochemical Parasitology*, 2007 156, 41-50.

[61] Siegel, TN; Kawahara, T; Degrasse, JA; Janzen, CJ; Horn, D; Cross, GA. Acetylation of histone H4K4 is cell cycle regulated and mediated by HAT3 in *Trypanosoma brucei*. *Molecular Microbiology*, 2008 67, 762-771.

[62] Aslund, L; Carlsson, L; Henriksson, J; Rydaker, M; Toro, GC; Galanti, N; Pettersson, U. A gene family encoding heterogeneous histone H1 proteins in *Trypanosoma cruzi*. *Molecular And Biochemical Parasitology*, 1994 65, 317-330.

[63] Solari, AJ. The 3-dimensional fine structure of the mitotic spindle in *Trypanosoma cruzi*. *Chromosoma*, 1980 78, 239-255.

[64] Nardelli, SC; da Cunha, JP; Motta, MC; Schenkman, S. Distinct acetylation of *Trypanosoma cruzi* histone H4 during cell cycle, parasite differentiation, and after DNA damage. *Chromosoma*, 2009 118, 487-499.

[65] Siegel, TN; Hekstra, DR; Kemp, LE; Figueiredo, LM; Lowell, JE; Fenyo, D; Wang, X; Dewell, S; Cross, GA. Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes & Development*, 2009 9 23 1063-1076.

[66] Talbert, PB; Henikoff, S. Chromatin-based transcriptional punctuation. *Genes & Development*, 2009 23, 1037-1041.

[67] Villanova. GV; Nardelli, SC; Cribb, P; Magdaleno, A; Silber, AM; Motta, MC; Schenkman, S; Serra, E. *Trypanosoma cruzi* bromodomain factor 2 (BDF2) binds to acetylated histones and is accumulated after UV irradiation. *International Journal For Parasitology*, 2009 39, 665-673.

For the exclusive use of R. Manning-Cela

[68] Ekanayake, D; Sabatini, R. Epigenetic regulation of polymerase II transcription initiation in *Trypanosoma cruzi*: modulation of nucleosome abundance, histone modification, and polymerase occupancy by O-linked thymine DNA glucosylation. *Eukaryotic Cell*, 2011 10, 1465-1472.

[69] Ekanayake, DK; Minning, T; Weatherly, B; Gunasekera, K; Nilsson, D; Tarleton, R; Ochsenreiter, T; Sabatini, R. Epigenetic regulation of transcription and virulence in *Trypanosoma cruzi* by O-linked thymine glucosylation of DNA. *Molecular And Cellular Biology*, 2011 31, 1690-1700.

[70] Cohen-Freue, G; Holzer, TR; Forney, JD; McMaster, WR. Global gene expression in *Leishmania*. *International Journal For Parasitology*, 2007 37, 1077-1086.

[71] Diehl, S; Diehl, F; El-Sayed, NM; Clayton, C; Hoheisel, JD. Analysis of stage-specific gene expression in the bloodstream and the procyclic form of *Trypanosoma brucei* using a genomic DNA-microarray. *Molecular And Biochemical Parasitology*, 2002 123, 115-123.

[72] Luu, VD; Brems, S; Hoheisel, JD; Burchmore, R; Guilbride, DL; Clayton, C. Functional analysis of *Trypanosoma brucei* PUF1. *Molecular And Biochemical Parasitology*, 2006 150, 340-349.

[73] Minning, TA; Weatherly, DB; Atwood, J3rd; Orlando, R; Tarleton, RL. The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. *BMC Genomics*, 2009 10, 370.

[74] Clayton, CE. Life without transcriptional control? From fly to man and back again. *The EMBO Journal*, 2002 21, 1881-1888.

[75] Martinez-Calvillo, S; Vizuet-de-Rueda, JC; Florencio-Martinez, LE; Manning-Cela, RG; Figueroa-Angulo, EE. Gene expression in trypanosomatid parasites. *Journal Of Biomedicine And Biotechnology*, 2010 2010, 525241.

[76] Benz, C; Nilsson, D; Andersson, B; Clayton, C; Guilbride, DL. Messenger RNA processing sites in *Trypanosoma brucei*. *Molecular And Biochemical Parasitology*, 2005 143, 125-134.

[77] Hummel, HS; Gillespie, RD; Swindle, J. Mutational analysis of 3' splice site selection during trans-splicing. *The Journal Of Biological Chemistry*, 2000 275, 35522-35531.

[78] Siegel. TN; Tan, KS; Cross, GA. Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. *Molecular And Cellular Biology*, 2005 25, 9586-9594.

[79] Liang, XH; Haritan, A; Uliel, S; Michaeli, S. Trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryotic Cell*, 2003 2, 830-840.

[80] Lopez-Estrano, C; Tschudi, C; Ullu, E. Exonic sequences in the 5' untranslated region of alpha-tubulin mRNA modulate trans splicing in *Trypanosoma brucei*. *Molecular And Cellular Biology*, 1998 18, 4620-4628.

[81] Huang, J; Van der Ploeg, LH. Requirement of a polypyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3' splice acceptor site. *The EMBO Journal*, 1991 10, 3877-3885.

[82] Matthews, KR; Tschudi, C; Ullu, E. A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes & Development*, 1994 8, 491-501.

[83] Manning-Cela, R; Gonzalez, A; Swindle, J. Alternative splicing of LYT1 transcripts in *Trypanosoma cruzi*. *Infection And Immunity*, 2002 70, 4726-4728.

For the exclusive use of R. Manning-Cela

[84] Benabdellah, K; Gonzalez-Rey, E; Gonzalez, A. Alternative trans-splicing of the *Trypanosoma cruzi* LYT1 gene transcript results in compartmental and functional switch for the encoded protein. *Molecular Microbiology*, 2007 65, 1559-1567.

[85] Kunz, S; Luginbuehl, E; Seebeck, T. Gene conversion transfers the GAF-A domain of phosphodiesterase TbrPDEB1 to one allele of TbrPDEB2 of *Trypanosoma brucei*. *PLoS Neglected Tropical Diseases*, 2009 3, e455.

[86] Manning-Cela, R; Cortes, A; Gonzalez-Rey, E; Van Voorhis, WC; Swindle, J; Gonzalez, A. LYT1 protein is required for efficient in vitro infection by *Trypanosoma cruzi*. *Infection And Immunity*, 2001 69, 3916-3923.

[87] Rinehart, J; Horn, EK; Wei, D; Soll, D; Schneider, A. Non-canonical eukaryotic glutaminyl- and glutamyl-tRNA synthetases form mitochondrial aminoacyl-tRNA in *Trypanosoma brucei*. *The Journal Of Biological Chemistry*, 2004 279, 1161-1166.

[88] Di Noia, JM; D'Orso, I; Sanchez, DO; Frasch, AC. AU-rich elements in the 3'-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. *The Journal Of Biological Chemistry*, 2000 275, 10218-10227.

[89] Hehl, A; Vassella, E; Braun, R; Roditi, I. A conserved stem-loop structure in the 3' untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*. *Proceedings Of The National Academy Of Sciences USA*, 1994 91, 370-374.

[90] Irmer, H; Clayton, C. Degradation of the unstable EP1 mRNA in *Trypanosoma brucei* involves initial destruction of the 3'-untranslated region. *Nucleic Acids Research*, 2001 29, 4707-4715.

[91] Batista, JA; Teixeira, SM; Donelson, JE; Kirchhoff, LV; de Sa, CM. Characterization of a *Trypanosoma cruzi* poly(A)-binding protein and its genes. *Molecular And Biochemical Parasitology*, 1994 67, 301-312.

[92] D'Orso, I; Frasch, AC. TcUBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes. *The Journal Of Biological Chemistry*, 2001 276, 34801-34809.

[93] Parodi-Talice, A; Monteiro-Goes, V; Arrambide, N; Avila, AR; Duran, R; Correa, A; Dallagiovanna, B; Cayota, A; Krieger, M; Goldenberg, S; Robello, C. Proteomic analysis of metacyclic trypomastigotes undergoing *Trypanosoma cruzi* metacyclogenesis. *Journal Of Mass Spectrometry*, 2007 42, 1422-1432.

[94] Ferella, M; Nilsson.D; Darban, H; Rodrigues, C; Bontempi EJ; Docampo, R; Andersson, B. Proteomics in *Trypanosoma cruzi*--localization of novel proteins to various organelles. *Proteomics*, 2008 8, 2735-2749.

[95] Cunha-e-Silva, NL; Atella, GC; Porto-Carreiro, IA; Morgado-Diaz, JA; Pereira, MG; de Souza, W. Isolation and characterization of a reservosome fraction from *Trypanosoma cruzi*. *FEMS Microbiology Letters*, 2002 214, 7-12.

[96] Sant'Anna, C; Nakayasu, ES; Pereira, MG; Lourenco, D; de Souza, W; Almeida, IC; Cunha, ESNL. Subcellular proteomics of *Trypanosoma cruzi* reservosomes. *Proteomics*, 2009 9, 1782-1794.

[97] Atwood, JA3rd; Minning, T; Ludolf, F; Nuccio, A; Weatherly, DB; Alvarez-Manilla, G; Tarleton, R; Orlando, R. Glycoproteomics of *Trypanosoma cruzi* trypomastigotes using subcellular fractionation, lectin affinity, and stable isotope labeling. *Journal Of Proteome Research*, 2006 5, 3376-3384.

For the exclusive use of R. Manning-Cela

[98] Nakayasu, ES; Yashunsky, DV; Nohara, LL; Torrecilhas, AC; Nikolaev, AV; Almeida, IC. GPIomics: global analysis of glycosylphosphatidylinositol-anchored molecules of *Trypanosoma cruzi*. *Molecular Systems Biology*, 2009 5, 261.

[99] Brenchley, R; Tariq, H; McElhinney, H; Szoor, B; Huxley-Jones, J; Stevens, R; Matthews, K; Tabernero, L. The TriTryp phosphatome: analysis of the protein phosphatase catalytic domains. *BMC Genomics*, 2007 8, 434.

[100] da Cunha, JP; Nakayasu, ES; Elias, MC; Pimenta, DC; Tellez-Inon, MC; Rojas, F; Munoz, MJ; Almeida, IC; Schenkman, S. *Trypanosoma cruzi* histone H1 is phosphorylated in a typical cyclin dependent kinase site accordingly to the cell cycle. *Molecular And Biochemical Parasitology*, 2005 140, 75-86.

[101] Bartholomeu, DC; Cerqueira, GC; Leao, AC; daRocha, WD; Pais, FS; Macedo, C; Djikeng, A; Teixeira, SM; El-Sayed, NM. Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. *Nucleic Acids Research*, 2009 37, 3407-3417.

[102] El-Sayed, NM; Myler, PJ; Blandin, G; Berriman, M; Crabtree, J; Aggarwal, G; Caler, E; Renauld, H; Worthey, EA. et al. Comparative genomics of trypanosomatid parasitic protozoa. *Science*, 2005 309, 404-409.

[103] Freitas, LM; dos Santos, SL; Rodrigues-Luiz, GF; Mendes, TA; Rodrigues, TS; Gazzinelli, RT; Teixeira, SM; Fujiwara, RT; Bartholomeu, DC. Genomic analyses, gene expression and antigenic profile of the trans-sialidase superfamily of *Trypanosoma cruzi* reveal an undetected level of complexity. *PLoS One*, 2011 6, e25914.

[104] Cremona, ML; Sanchez, DO; Frasch, AC; Campetella, O. A single tyrosine differentiates active and inactive *Trypanosoma cruzi* trans-sialidases. *Gene*, 1995 160, 123-128.

[105] Schenkman, S; Eichinger, D; Pereira, ME; Nussenzweig. V. Structural and functional properties of *Trypanosoma* trans-sialidase. *Annu Rev Microbiol*, 1994 48, 499-523.

[106] Buscaglia, CA; Campo, VA; Di Noia, JM; Torrecilhas, AC; de Marchi, CR; Ferguson, MA; Frasch, AC; Almeida, IC. The surface coat of the mammal-dwelling infective trypomastigote stage of *Trypanosoma cruzi* is formed by highly diverse immunogenic mucins. *The Journal Of Biological Chemistry*, 2004 279, 15860-15869.

[107] Campo, V; Di Noia, JM; Buscaglia, CA; Aguero, F; Sanchez, DO; Frasch, AC. Differential accumulation of mutations localized in particular domains of the mucin genes expressed in the vertebrate host stage of *Trypanosoma cruzi*. *Molecular And Biochemical Parasitology*, 2004 133, 81-91.

[108] Kawashita, SY; da Silva, CV; Mortara, RA; Burleigh, BA; Briones, MR. Homology, paralogy and function of DGF-1, a highly dispersed *Trypanosoma cruzi* specific gene family and its implications for information entropy of its encoded proteins. *Molecular And Biochemical Parasitology*, 2009 165, 19-31.

For the exclusive use of R. Manning-Cela

In: Comparative Genomics in Neglected Human Parasites
ISBN: 978-1-62417-874-0
Editors: César López-Camarillo and L. A. Marchat
© 2013 Nova Science Publishers, Inc.

Chapter 4

COMPARATIVE GENOMICS OF *LEISHMANIA* PARASITES

Norma E. Padilla-Mejía[a,b], Claudia M. Gómez-Hurtado[a], Inti I. Sánchez-Santamaría[a], Luis E. Florencio-Martínez[a], Rebeca G. Manning-Cela[c] and Santiago Martínez-Calvillo*[a,b]

[a]Unidad de Biomedicina, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Col. Los Reyes Iztacala, Tlalnepantla, Edo. de México, Mexico
[b]Posgrado en Ciencias Biológicas, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Col. Los Reyes Iztacala, Tlalnepantla, Edo. de México, Mexico
[c]Departamento de Biomedicina Molecular, Centro de Investigación y de Estudios Avanzados del IPN, Col. San Pedro, Zacatenco, México, D.F., Mexico

ABSTRACT

The parasites of the genus *Leishmania* are trypanosomatid protozoa that produce a spectrum of diseases ranging from mild skin or mucosal lesions to fatal visceral leishmaniasis. The genome sequences of five different species of *Leishmania* have been reported to date: *L. major*, *L. infantum*, *L. braziliensis*, *L. mexicana* and *L. donovani*. Analyses of the sequences revealed that the genomes of these parasites are organized into large directional gene clusters, *i.e.* tens-to-hundreds of protein-coding genes arranged sequentially on the same strand of DNA. A remarkable conservation of gene order (synteny) was observed in the genomes of the different *Leishmania* species. Interestingly, an unexpectedly small number of species-specific genes was identified. However, the analyses showed gene and chromosome copy number differences between species, indicating that increased gene copy number may cause changes in gene expression that might influence disease tropism. Contrary to what occurs in other eukaryotes, transcription in *Leishmania* and other trypanosomatids is polycistronic, and mature messenger RNAs (mRNAs) are generated from primary transcripts by *trans*-splicing and

* Corresponding author: Email: scalv@campus.iztacala.unam.mx, Phone: (+52) (55) 5623-1333 ext. 39782, Fax: 00 (+52) (55) 5623-1138.

For the exclusive use of R. Manning-Cela

polyadenylation. Little is known about either the DNA sequences or the proteins that are involved in transcription initiation in *Leishmania*. Bioinformatic analyses of the genome databases of these parasites led to the identification of a small number of proteins involved in gene expression. However, functional studies have revealed that trypanosomatids have more general transcription factors than originally estimated. Analysis of the *Leishmania* databases showed the presence of a relatively low number of transfer RNA (tRNA) genes, which are organized into clusters of two to 10 genes that may contain other Pol III-transcribed genes. Also, with the exception of *L. braziliensis*, the *Leishmania* genome does not contain active retrotransposons. Interestingly, *L. braziliensis* possesses components of an RNA-mediated interference pathway, which is not present in other *Leishmania* species.

1. Introduction

The members of the Trypanosomatidae family (order Kinetoplastida) are flagellated protozoa of both biological and medical importance. This family includes parasites of the genus *Leishmania* and *Trypanosoma*, which exhibit complex life cycles with several developmental stages that alternate between vertebrate and invertebrate hosts. *Leishmania* is the causative agent of leishmaniasis, a disease with a wide range of clinical manifestations that arise from infections with different species of the parasite [1]. There are three main clinical forms of leishmaniasis: cutaneous, mucocutaneous and visceral [2; 3]. Cutaneous leishmaniasis is the most common form and it typically produces ulcers, which heal spontaneously, on the face and arms of the infected person. However, ulcers cause serious disability and leave severe and permanently disfiguring scars. Mucocutaneous leishmaniasis is the most disfiguring clinical form of the disease, since it causes extensive destruction of the oral-nasal and pharyngeal cavities with dreadful disfiguring lesions and mutilation of the face. Visceral leishmaniasis is characterized by fever, substantial weight loss and swelling of the spleen and liver, and it is fatal in the absence of treatment [4]. *Leishmania* parasites are transmitted to humans and other vertebrates through the bites of infected sandflies of the genus *Lutzomyia* (in the New World) and *Phlebotomus* (in the Old World). The metacyclic promastigote, the infective form of the parasite, invades the macrophages and differentiates into amastigotes, which are the proliferative forms within the vertebrate host. The parasite replicates within the insect vector as a non-infective procyclic promastigote [5]. According to the World Health Organization (WHO), leishmaniasis is endemic in 98 countries or territories in tropical and sub-tropical areas of the world [6]. Approximately 350 million people are considered at risk of contracting leishmaniasis, and some two million new cases occur yearly. Visceral leishmaniasis causes an estimated over 50,000 deaths annually, a rate surpassed among parasitic diseases only by malaria. Thus, the *Leishmania* parasites represent an important global health problem for which there is no vaccine and few effective medicines. The recent publication of the genome sequences of several species of *Leishmania* will not only help to improve vaccine and drug design, but also to understand diverse molecular aspects and the evolution of the parasite.

For the exclusive use of R. Manning-Cela

2. Diversity of *Leishmania* Species

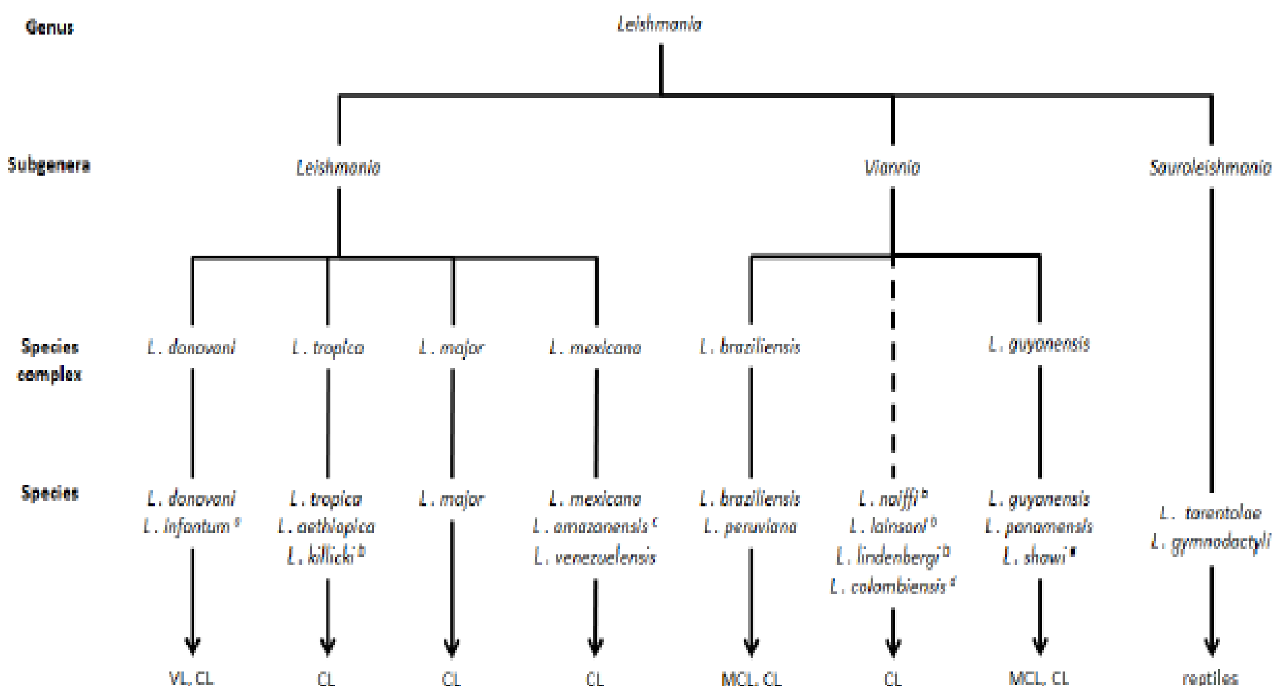
In 1903, William Leishman and Charles Donovan independently described *Leishmania* parasites, but they were previously observed by David D. Cuningham in 1885 and Peter Borovsky in 1898 [7]. The parasite was first cultivated in 1904 by Leonard Rogers [8]. To date, more than 30 different species of *Leishmania* have been described in five different continents, and at least 20 of them cause disease in human [6]. Molecular methodologies allow accurate species identification. Isoenzyme electrophoresis is currently the reference identification technique, and correlations have been established between clinical forms and zymodemes (parasite populations with common isoenzyme patterns) for some species. Kinetoplast DNA (kDNA) digested with restriction enzymes has also helped to identify species and parasite populations (schizodemes). Other DNA techniques, more sensitive and easier to use, will probably prevail in the future.

Depending on which parts of the sandfly gut are colonized by the parasite, the genus *Leishmania* is divided into three subgenera: *Leishmania*, *Viannia* and *Sauroleishmania* [9-11]. Species classified within the *Leishmania* subgenus produce cutaneous and visceral leishmaniasis in both the New World and the Old World, while species included in the *Viannia* subgenus cause cutaneous and mucocutaneous leishmaniasis only in the New World. Species that belong to the *Sauroleishmania* subgenus infect reptiles and are not pathogenic to humans. The three *Leishmania* subgenera are divided into several complexes that include different species (figure 1).

Leishmania and the rest of the trypanosomatids diverged early from the main eukaryotic lineage. As a consequence, this group of organisms shows numerous unusual features at the genetic, biochemical and cytological level [5; 12; 13]. For example, *Leishmania* presents a single branched mitochondrion that extends the whole length of the cell and can fill up to 12% of the cell volume. The mitochondria contain the kDNA, which is the most unusual structure in the organelle (see below). *Leishmania* extensively decodes its mitochondrial transcripts through RNA editing, a process that involves specific uridine insertion and deletion that can double the size of the primary transcript, and that is essential for creation of translatable open reading frames [14; 15].

Another organelle that is exclusive to *Leishmania* and other trypanosomatids is the glycosome, a peroxisome-related structure that contains enzymes involved in several metabolic pathways, including glycolysis. Glycosomes are bounded by a single membrane and have a protein-dense matrix [16]. Another unique characteristic of trypanosomatids is the presence of subpellicular microtubules, which is a layer of microtubules that run longitudinally underneath the plasma membrane [17]. Remodelling of this cytoskeleton facilitates the distinct morphologies of the various developmental forms of *Leishmania*. In trypanosomatids, the flagellum consists of a 9+2 microtubule configuration axoneme and a paraflagellar rod, a unique lattice-like structure that lies alongside the axoneme. The flagellum emerges from a flagellar pocket, a small invagination of the plasma membrane that constitutes the only site of exocytosis and endocytosis in the cell [18].

For the exclusive use of R. Manning-Cela



VL, visceral leishmaniasis; CL, cutaneous leishmaniasis; MCL, mucocutaneous leishmaniasis

[a] In the New World, this species is known as *L. chagasi*
[b] The taxonomic status of this species is under discussion.
[c] *L. amazonensis* can cause also MCL
[d] *L. colombiensis* can cause also VL. Species taxonomic status is under discussion
[e] *L. shawi* seems to cause only CL

Figure 1. Taxonomy of *Leishmania*. The different species are organized into three subgenera and several complexes. The clinical forms of leishmaniasis produced by the different species are indicated. Members of the *Sauroleishmania* subgenus infect reptiles only.

For the exclusive use of R. Manning-Cela

3. Genome Organization

The *Leishmania* genome is organized into large directional gene clusters (DGCs), *i.e.* tens-to-hundreds of protein-coding genes arranged sequentially on the same strand of DNA. This striking gene organization was initially observed on *L. major* Friedlin chromosome 1, the first entirely sequenced chromosome in trypanosomatids, which contains 85 genes organized into two divergent DGCs, with the first 32 genes clustered on the bottom strand and the remaining 53 genes grouped on the top strand [19]. The publication of the complete genome of *L. major* Friedlin revealed that the genes in all the chromosomes are organized into large DGCs [20]. A similar gene organization was reported for *T. brucei* [21] and *T. cruzi* [22], indicating that the presence of DGCs is another unusual feature that is present in all trypanosomatids. Recently, the genome sequences of other four species of *Leishmania* were reported: *L. infantum* (strain JPCM5), *L. braziliensis* (M2904), *L. mexicana* (U1103) and *L. donovani* (BPK282/0cl4) [23-25]. These reports provided resources for comparative genomic studies, and information about the genes that might have a role in the tissue-specific development of the different species of *Leishmania*.

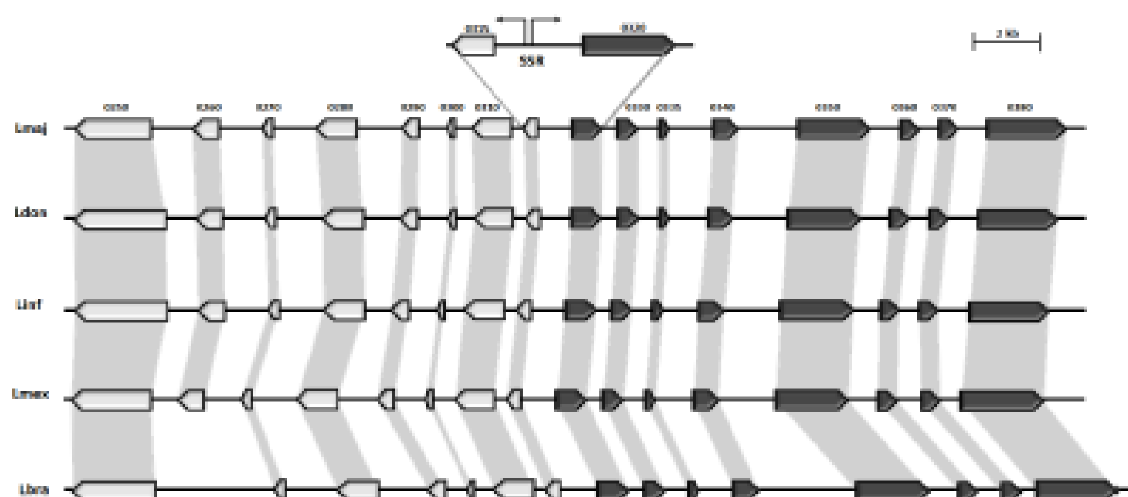


Figure 2. Synteny of protein-coding genes on chromosome 1 from *Leishmania*. The first eight genes of each of the two directional gene clusters present on this chromosome are shown for *L. major* (Lmaj), *L. donovani* (Ldon), *L. infantum* (Linf), *L. mexicana* (Lmex), and *L. braziliensis* (Lbra). The names of the genes, indicated only for *L. major*, are abbreviated (for instance, 0250 corresponds to LmjF01.0250). Orthologous genes are joined by grey lines. In the enlargement of the strand switch region (SSR) from *L. major*, the transcription start sites for both DGCs are denoted with arrows.

The genomes of the different species of *Leishmania* show a noteworthy conservation of gene order (synteny), despite an estimated divergence of 46 million years [26]. Gene order is conserved for more than 99% of genes between *L. major, L. infantum* and *L. braziliensis* [23]. To exemplify synteny, figure 2 shows a schematic representation of the strand switch region (SSR) and the first eight genes of the two DGCs present on chromosome 1 from several species of *Leishmania*. As observed, the order of the genes is exactly the same in all the species. The only exception is the orthologue of gene 0260 (LmjF01.0260), which was lost in

For the exclusive use of R. Manning-Cela

the *L. braziliensis* genome (figure 2). Sequence conservation of coding regions is high among different species of *Leishmania*; for example, the average amino acid identity between *L. major* and *L. infantum* is 92%, and the average nucleotide identity is 94% [23].

Table 1. Summary of *Leishmania, T. brucei* and *T. cruzi* genomes.

	L. major	*L. infantum*	*L. braziliensis*	*L. mexicana* [24]*	*L. donovani* [25]	*T. brucei* [22]	*T. cruzi* [21]
Chromosome number	36 [20]	36 [24]	35 [26]	34	36	11 a	41 [27]
Genome size (bp)	32,855,089 [24]	32,101,728 [24]	31,997,773 [24]	32,108,741	32,444,998	26,075,396	~60,000,000
Overall G+C content (%)	59.7 [24]	59.3 [26]	57.8 [24]	59.7	NR	46.4	51
Coding G+C content (%)	62.5 [23]	62.45 [23]	60.38 [23]	NR c	NR	50.9	53.4
Predicted protein-coding genes	8,412 [24]	8,214 [26]	8,357 [26]	8,250	NR	9,068	~12,000
Pseudogenes	97 [23]	41 [23]	161 [23]	99	13	904	3,590
Intergenic regions mean lengh (bp)	2,045 [20]	2,049 [26]	1,976 [26]	NR	NR	1,279	1,024
Gene density (genes/Mb)	252 [20]	235 [26]	258 [26]	NR	NR	317	385
tRNA genes	83 [31]	77 b	75 b	84 b	76 b	66 [31]	120 [31]

a) Only mega chromosomes
b) These numbers correspond to the genes annotated in GeneDB (*L infantum*, 67 genes; *L. brasiliensis* 66; *L. mexicana*, 83, *L. donovani*, 64), plus the genes that we found using tRNAScan SE
c) NR: not reported
*Reference numbers are shown in brackets

The size of the haploid genomes in the diverse species of *Leishmania* is very similar, ranging from 31.9 to 32.8 megabases (Mb) in *L. brasiliensis* and *L. major*, respectively (table 1). These genomes are slightly larger than the one from *T. brucei* (26 Mb), but smaller than the one from *T. crusi* (~60 Mb). The number of predicted protein-coding genes is 8412 in *L. major*, 8241 in *L. infantum*, 8357 in *L. braziliensis* and 8250 in *L. mexicana* [24]. The estimated number of protein-coding genes is 9068 in *T. brucei* [22] and ~12000 in *T. cruzi* [27]. The majority of the predicted genes (~70%) encode hypothetical proteins of unknown function. These may represent genes that have parasite-specific functions, or which are sufficiently diverged as to have no significant sequence similarity to their functional homologs in other species [28].

In the *Leishmania* genomes, most protein-coding genes are separated by relatively short intergenic regions (mean length of 2045 bp in *L. major*). The density of protein-coding genes in the *L. major* genome is 252 genes per Mb, which is similar to what has been found in other trypanosomatids (317 genes per Mb in *T. brucei* and 385 genes per Mb in *T. cruzi*), but considerably higher than what has been reported in human (~8 genes per Mb) [29]. The overall G+C content of the genomes of the various species of *Leishmania* (from 57.8 to 59.7%) is higher than G+C content in *T. brucei* (46.4%) and *T. cruzi* (51%), and the G+C content of the coding regions is even higher (62.5% in *L. major*) (table 1). Similarly to other eukaryotes, the ends of the chromosomes in *Leishmania* contain the telomeric repeat

For the exclusive use of R. Manning-Cela

GGGTTA. Chromosomes contain short subtelomeric regions composed of variable repetitive elements that are responsible for a major part of the polymorphisms observed between homologous chromosomes. In contrast, subtelomeric regions in *T. brucei* are long and encode species-specific genes. DNA repeats represent 9-10% of *Leishmania* genome. Also, conserved amino acid repeats, which have been hypothesized to have a role in pathogenicity, are present in 3-4% of the predicted proteins in *Leishmania* [24]. The vast majority of protein-coding genes in trypanosomatids lack introns; in fact, *cis*-splicing has only been demonstrated for the gene encoding the poly-A polymerase [30].

Trypanosomatids do not seem to contain DNA transposons. However, analysis of the *T. cruzi* and *T. brucei* genomes confirmed the presence of abundant long terminal repeat (LTR) and non-LTR retrotransposons [21; 31]. They account for ~5% and 2% of the *T. cruzi* and *T. brucei* genomes, respectively. In contrast, *L. major and L. infantum* do not contain active retrotransposons. Nevertheless, they have remnants of extinct ingi/L1*Tc*-like retroposons called DIREs [32]. Recently, two new families of degenerated retrotransposons were identified in *Leishmania*: SIDER1 and SIDER2 [33; 34]. These sequences are predominantly located within the 3´-UTR of *Leishmania* mRNAs, and it was shown that SIDER2 acts as an instability element, since SIDER2-containing mRNAs are generally expressed at lower levels compared to the non-SIDER2 mRNAs [33].

Surprisingly, recent findings showed the presence of potentially active SLACS/CZAR retroposons in the *L. braziliensis* genome [23; 35]. These elements are associated with tandemly repeated spliced leader genes, showing an arrangement similar to that of the SLACS or CZAR elements in *T. brucei* or *T. cruzi*, respectively. Moreover, the telomeric regions of *L. braziliensis* contain a family of 20-30 previously unknown transposable elements called "telomere-associated transposable elements" (TATEs). Each TATE includes putative reverse transcriptase, phage integrase, and DNA and/or RNA polymerase sequences. TATEs are present on at least 12 chromosomes, inserted in the middle of the GGGTTA telomeric repeats [36].

4. Species-Specific Genes and RNA Interference

Sequence comparisons among the different species of *Leishmania* showed a surprisingly small number of species-specific genes, in spite of the broad differences in pathogenicity and immune response [23; 24]. In *L. mexicana* it was reported the presence of only two unique genes, which encode predicted proteins of unknown function. Interestingly, one of these genes is predicted to be a pseudogene in other *Leishmania* species, but it is the orthologue of an intact gene in *T. brucei*. *L. major* possesses 14 specific genes: 13 encode proteins of unknown function and one encodes a PFPI-like peptidase. In *L. braziliensis* there are 67 unique genes, of which 54 are predicted to encode proteins of unknown function. Species from the *L. donovani* complex (*L. donovani* and *L. infantum*) contain 19 specific genes, of which 15 encode proteins of unknown function; these genes might play an important role in the potential for visceralization associated with the *L. donovani* complex [25]. The location of the species-specific genes in *Leishmania* is also unexpected, since they are dispersed all over the genome rather than clustered in subtelomeric regions or breakpoints of DGCs, as reported in other trypanosomatids [22].

For the exclusive use of R. Manning-Cela

To determine if species-specific genes are present in different strains of the same species, second strains of *L. major* and *L. mexicana* were sequenced [24]. It was found that the second strain of *L. mexicana* (M379) contains intact copies of both unique genes identified in *L. mexicana* U1103. Similarly, the strain LV39 from *L. major* contains 13 of the 14 specific genes found in *L. major* Friedlin. Moreover, 17 different strains of *L. donovani* isolated from patients from India, Nepal and Ethiopia contained the 19 unique genes found in the *L. donovani* complex [25]. Thus, specific genes in *Leishmania* are conserved between strains of the same species, even when they were isolated from different geographical regions. However, extensive variation in gene and chromosome copy number was reported among different species and strains of *Leishmania* (see below).

In *Leishmania*, establishment of pseudogenes and gene loss seem to be the main factors that influence the evolutionary formation of new species. The species specificity of most genes can be attributed to the deterioration of a gene in the other species. In other words, in the species that lack the functional gene there is a degenerate sequence (pseudogene) in the corresponding region of synteny [23; 24]. For example, one of the unique genes in *L. infantum*, the one that encodes a putative phosphatidylinositol 3-kinase (PI3K), is predicted to encode a pseudogene in the other species of *Leishmania*. Also, A2 is one of the *L. donovani*-specific genes that has been proposed to be required for parasite survival in visceral organs. *L. major* and *L. tropica* contain only A2 pseudogenes [37].

RNA interference (RNAi) is a mechanism of gene silencing present in many eukaryotes. In this process, short double-stranded RNA molecules induce the sequence-specific degradation of homologous mRNA. It is thought that RNAi is not only involved in the control of endogenous mRNAs, but also has a defensive role against transposon activity and viruses [38]. Numerous proteins participate in RNAi, including Dicer and Argonaute. Several years ago a functional RNAi pathway was identified in *T. brucei* [39]. In contrast, *T. cruzi*, *L. major* and *L. donovani* lack Dicer and Argonaute activities and associated genes [40]. These findings led to the assumption that *T. brucei* was the only trypanosomatid with active RNAi mechanisms. Consequently, it was surprising when genes involved in the RNAi pathway were identified in *L. braziliensis* [23; 35] and other species of the *Viannia* subgenus, such as *L. guyanensis* and *L. panamensis* [41]. It was later demonstrated that in fact *L. braziliensis* shows strong RNAi activity with reporter and endogenous genes [41]. Functional and evolutionary studies of RNAi genes established that RNAi must have been lost twice independently in trypanosomatids: once in the lineage leading to *T. cruzi*, and a second time after the separation of the *Viannia* subgenus from the remaining *Leishmania* species [41]. In *L. braziliensis*, the Argonaute gene is located on chromosome 11. Sequence analysis of the syntenic regions in *L. major* and *L. infantum* revealed highly degenerate Argonaute pseudogenes, which supports the hypothesis that RNAi-related genes were lost in members of the *Leishmania* subgenus [36].

5. Gene and Chromosome Copy Number Variation

Importantly, differences in gene content and chromosome copy number between species and strains of *Leishmania* have been recognized as a major source of genomic variation. Regarding chromosome architecture, *L. major*, *L. donovani* and *L. infantum* have 36

For the exclusive use of R. Manning-Cela

chromosomes, whereas *L. braziliensis* has only 35 chromosomes due to the fusion of chromosomes 20 and 34. The genome of *L. mexicana* is organized into 34 chromosomes, as the result of two fusion events involving chromosomes 8 and 29, and chromosomes 20 and 36 [42]. *Leishmania* is generally considered to be diploid, although several chromosomes have been described as aneuploid [43; 44]. Homologous chromosomes can vary in size, as it has been described for the homologous copies of chromosome 1 in *L. major* Friedlin, which differ in size by approximately ~30 kb [45].

Chromosome copy number has been analyzed for different species of *Leishmania* by chromosome Read Depth and other methodologies such as allele frequency distribution plots, FISH and FACS [24; 25; 46]. Analysis of two *L. major* strains, Friedlin and LV39, showed that most chromosomes in this species are disomic, except for chromosome 31 (reported as tetrasomic), chromosome 1 (trisomic) and chromosome 34 (reported as supernumerary). Several trisomic chromosomes have been reported in *L. infantum* JPCM5 (nine chromosomes), *L. donovani* LV9 (three chromosomes) and *L. mexicana* U1103 (three chromosomes). In *L. donovani* BPK206/0, five chromosomes are trisomic and two chromosomes are tetrasomic. In different species, several chromosomes appear to be of intermediate status, being neither disomic nor trisomic. These patterns may result from a mixture of individual cells within a population with monosomic, disomic and trisomic chromosomes. Interestingly, in the strain M2904 of *L. braziliensis* 30 of 35 chromosomes are clearly trisomic, three are tretrasomic and one is hexasomic (chromosome 31). Thus, in contrast to other species of *Leishmania*, *L. braziliensis* M2904 is primarily triploid [24]. It is worth mentioning that chromosome 31 is the only one that is supernumerary in all the species and isolates of *Leishmania* analyzed, including the homologous chromosome 30 in *L. mexicana*. The functional meaning of this finding has yet to be determined. In conclusion, *Leishmania* parasites can contain disomic, trisomic, tetrasomic or supernumerary chromosomes, depending on the species, and the pattern of aneuploidy seems to be characteristic for a specific strain.

It is also important to mention that multicopy gene arrays are present more frequently in disomic chromosomes than in supernumerary chromosomes in four species of *Leishmania* analyzed (*L. braziliensis, L. major, L. mexicana* and *L. infantum*). It has been suggested that these disomic chromosomes may have persisted as non-supernumerary because of fitness constraints that are gene-dosage related. Since gene-dosage sensitivity is known to be important in selection of copy number changes, in the *Leishmania* species there may be selection against whole-chromosome duplications for those chromosomes that have a higher proportion of dose-sensitive genes [24].

Analysis of allelic variation among different *Leishmania* species by heterozygous single-nucleotide polymorphisms (SNPs) showed a remarkably low level of heterozygosity in *L. major* (297 SNPs) and *L. infantum* (629 SNPs) compared with *L. mexicana* (12,531 SNPs) and *L. braziliensis* (44,588 SNPs). This finding could result from a higher degree of inbreeding within *L. major* and *L. infantum* relative to *L. braziliensis* and *L. mexicana*. The high level of heterozygosity observed in *L. braziliensis* could in some measure be attributable to its triploidy and the supernumeracy nature of many of its chromosomes, since redundancy of essential genes may allow a higher rate of neutral mutation than that of the diploid genomes of *L. major* and *L. infantum* [24].

Among different strains of *L. donovani*, 3549 SNPs were identified [25]. About 82% of them were located in non-coding regions, 7% were synonymous mutations in coding regions

For the exclusive use of R. Manning-Cela

and 11% caused changes at the protein sequence level. Remarkably, some SNPs seem to be related to resistance to the antimonial drug sodium stibogluconate (SSG). Also, it is interesting that SSG-resistant parasites contain on average 44% fewer rDNA transcription units and 77% more mini-exon units than sensitive parasites. Moreover, *L. donovani* SSG-resistant lines show more copies of an episome that contains the mitogen-activated protein kinase (MAPK) locus.

6. Major Gene Families

Another characteristic feature of the *Leishmania* genome is the presence of tandem arrays of duplicated genes, which facilitate increased protein expression in the absence of the regulated transcriptional control that is found in other organisms. The number of tandem arrays differs among species: 132 in *L. mexicana* U1103, 200 in *L. major* Friedlin, 207 in *L. infantum* JPCM5 and 214 in *L. braziliensis* [24]. The number of genes per tandem array also varies among *Leishmania* species.

From the total number of protein-coding genes, only 56 were present as multicopy genes in all species of *Leishmania* [24]. These include the amastin genes, which encode small surface glycoproteins of unknown function that are also present in *T. cruzi*. The majority of amastin genes is expressed in intracellular amastigotes, and may contribute to their survival in the human host. In *L. major*, amastin genes are distributed on seven different chromosomes. The largest tandem array, located on chromosome 34, contains 23 amastin genes that alternate with the highly conserved tuzin genes (that encode proteins of unknown function) [47]. *L. infantum* and *L. braziliensis* have amastin gene arrays at the same locations; however, the large amastin-tuzin gene array seems to be shorter in *L. braziliensis* [23]. The *Leishmania* glycoprotein GP63, also known as leishmanolysin or major surface protease (MSP), is a surface metalloprotease involved in parasite survival, infectivity and virulence [48]. In *L. major*, it is encoded by a tandem array of seven genes located on chromosome 10. Interestingly, GP63 tandem array is composed of ~28 genes in *L. braziliensis* [23].

PSA, also called PSA-2 or GP46, is a *Leishmania* family of membrane-bound or secreted proteins, whose main signature consists in a specific array of leucine-rich repeats flanked by conserved cysteine-rich domains. It is overexpressed in metacyclic promastigotes, where it is involved in resistance to complement-mediated lysis. In *L. major*, 32 PSA genes have been identified on chromosomes 5, 9, 12, 21 and 31. Similarly, 14 and 8 PSA genes were found on the same chromosomes in *L. infantum* and *L. braziliensis*, respectively. The PSA subfamily present on chromosome 12 is the most complex in *L. major* and *L. infantum*, since it is composed of tandem arrays of 24 genes in the former and seven genes in the latter. In contrast, in *L. braziliensis* there is only one PSA gene on this locus [49]. Other multicopy genes present in the *Leishmania* genome include alpha- and beta-tubulin, HSP70, HSP83, kinesins, several translation initiation and elongation factors, protein kinases, RNA helicases, protein phosphatases, amino acid permeases and several hypothetical proteins [20].

For the exclusive use of R. Manning-Cela

7. Gene Expression in *Leishmania*

Eukaryotic cells usually have three distinct classes of nuclear RNA polymerases (Pol): Pol I, II, and III. Each class of polymerase transcribes a different kind of RNA. Pol I synthesizes 18S, 5.8S and 28S ribosomal RNAs (rRNAs), and Pol II transcribes mRNAs and most of the small nuclear RNAs (snRNAs). Pol III is involved in the production of small essential RNAs, such as tRNAs, 5S rRNA and some snRNAs.

Contrary to what occurs in other eukaryotes, Pol II transcription in *Leishmania* and other trypanosomatids is polycistronic [50-52]. Most chromosomes contain at least two PGCs, which can be either divergently transcribed (towards the telomeres) or convergently transcribed (away from the telomeres). Mature nuclear mRNAs are generated from primary transcripts by *trans*-splicing and polyadenylation (figure 3) [53; 54]. *Trans*-splicing is a process that adds a capped 39-nucleotide miniexon or spliced leader (SL) to the 5′ termini of the mRNAs [55; 56]. All the genes that are part of a PGC are transcribed at the same level, as a consequence of polycistronic transcription. However, the mature mRNAs of adjacent genes might show very different concentrations and/or stage-specific expression. This is because gene expression in trypanosomatids is mainly regulated posttranscriptionally at the level of mRNA processing and stability [13; 57]. Sequences in the 3′ untranslated region (3′-UTR) of an mRNA play a key role in gene expression. For example, the 3′-UTR from the amastin mRNA in *L. infantum* has a 450-bp region that confers amastigote-specific gene expression by a mechanism that increases the mRNA translation [58].

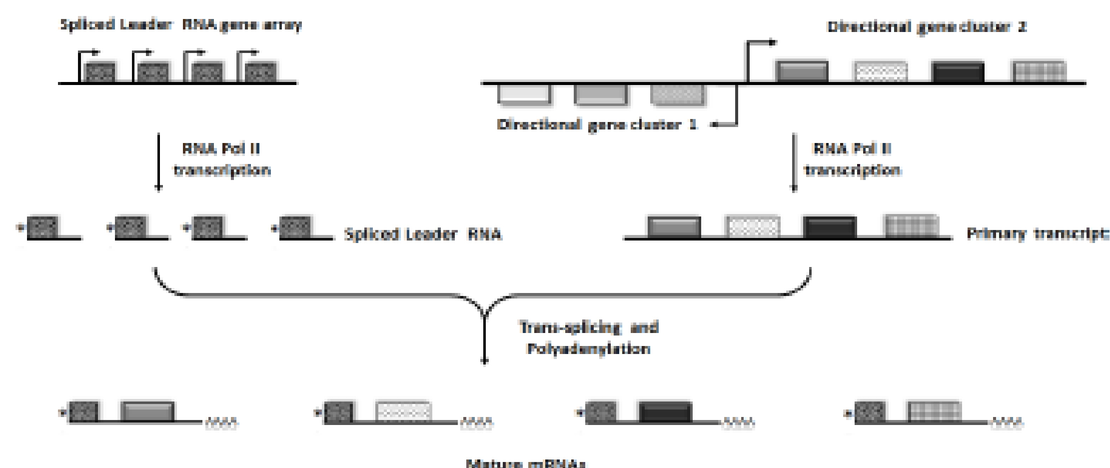


Figure 3. Transcription and processing of mRNAs in *Leishmania*. Pol II transcription initiates upstream of the first gene of the DGCs (arrows), generating primary transcripts (shown only for DGC2) that are processed by *trans*-splicing and polyadenylation to generate the mature mRNAs. By *trans*-splicing, a capped spliced leader RNA is added to the 5′ end of every mRNA. In the spliced leader locus (located on a different chromosome) each gene possesses a Pol II promoter region (arrows). The cap in the spliced leader RNA is indicated with an asterisk at the 5′ end of the RNA. The four As located at the 3′ end of the mature mRNAs represent the poly-A tail.

For the exclusive use of R. Manning-Cela

In trypanosomatids, identification of Pol II promoters for protein-coding genes has proven to be a difficult goal, complicated by a relatively low Pol II transcriptional activity and rapid processing of the primary transcripts. Nevertheless, transcriptional analysis of chromosome 1 from *L. major* showed that Pol II transcription of the entire chromosome initiates in the strand-switch region (between the two divergent PGCs) and proceeds bidirectionally towards the telomeres (figure 2) [52]. Several transcription start sites were mapped for both PGCs within a <100-bp region that contains long G-tracts (or C-tracts), but do not contain a TATA box or any other typical Pol II core promoter elements. Thus, while in most eukaryotes each gene possesses its own promoter, a single region seems to drive the expression of the entire chromosome 1 in *L. major* [52]. Similar studies performed on chromosome 3 from *L. major* confirmed that Pol II transcription initiates only upstream of the first gene of a PGC [59]. Since most genes are organized into large PGCs in trypanosomatids, the number of regions where transcription of Pol II initiates in these organisms is very low (only a few per chromosome) compared to other eukaryotes. There is no substantial sequence homology among the transcription initiation regions on *L. major* chromosomes 1 and 3. Hence, rigorously conserved sequence recognition sites do not appear to be required for Pol II transcription initiation in *Leishmania*. Interestingly, a recent ChIP-chip study in *L. major* showed that H3 histones acetylated at K9/K14, a marker for sites of active transcription initiation in other eukaryotes, are found at all divergent strand-switch regions in the parasite [60]. Moreover, peaks for two transcription factors, TBP and SNAP50, were also associated with divergent strand-switch regions [60].

In trypanosomatids, the only Pol II promoter that has been extensively characterized is the one driving the expression of the SL RNA [61-63]. In *L. tarentolae* it consists of two domains: the -60 element (from -67 to -58, relative to the TSS) and the -30 element (from -41 to -31). In trypanosomatids, transcription factors could not be identified by standard *in silico* analysis because they are extremely divergent [64; 65]. Consequently, their identification relied on biochemical, structural and functional analyses. Regarding Pol II transcription, several general transcription factors that participate in SL RNA synthesis have been identified in *T. brucei*. These include TBP, TFIIB, SNAPc, TFIIA and TFIIH [66-70]. Thus, these recent findings indicate that trypanosomatids possess more general transcription factors than initially estimated from *in silico* studies.

8. Genomic Organization of Transfer RNA and Ribosomal RNA Genes

Analysis of the *L. major* genome databases showed the presence of 83 tRNA genes distributed among 31 loci, on 19 different chromosomes [20; 71]. Most tRNA genes are organized into clusters of two to 10 genes, on either top or bottom strand, which may contain other Pol III-transcribed genes. For example, in the locus located on chromosome 23 there are 10 tRNA genes, a 5S rRNA gene and the U1 and U3 snRNA genes (figure 4). In *T. brucei*, a total of 66 tRNA genes were located on 26 loci, on eight different chromosomes. As in *L. major*, in *T. brucei* the number of tRNA genes per cluster ranges from two to 10. *In T. cruzi*, 120 tRNA genes were identified [71]. Thus, the number of tRNA genes in trypanosomatids is relatively low, considering that eukaryotic organisms usually contain several hundred tRNA

For the exclusive use of R. Manning-Cela

genes. For instance, *C. elegans* has 568 tRNA genes and *Homo sapiens* presents 497 tRNA genes [72; 73].

In contrast to protein-coding genes, the majority of the tRNA clusters do not show synteny in trypanosomatids, but a few of them do show conservation [71]. Among the latter, the most outstanding example is the cluster of 13 Pol III genes, located on chromosome 23 in *L. major*, that is highly syntenic. Interestingly, the order of the genes in this cluster is identical between *T. brucei* and *T. cruzi*, although three genes are located on different strands [71]. The majority of the 13 genes is present in the *L. major* cluster, but their order is not identical to either of the other two clusters (figure 4, compare *L. major* with *T. brucei*). Additionally, a 5S rRNA gene replaced a 7SL RNA gene and a tRNA-Trp gene replaced one of the tRNA-Lys genes.

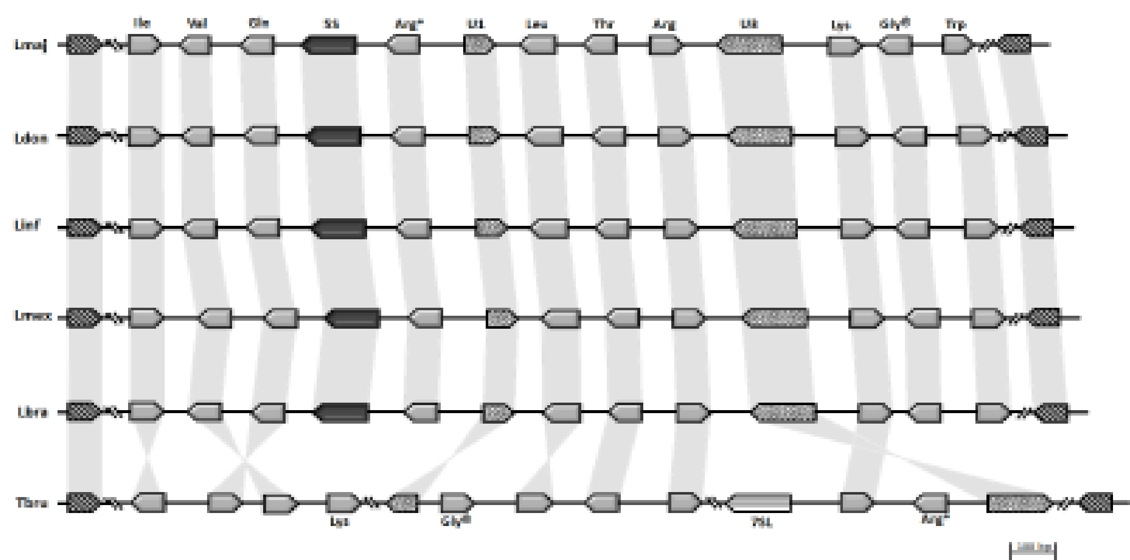


Figure 4. Comparative order of Pol III-transcribed genes in *Leishmania* and *T. brucei*. The order of the 13 Pol III-transcribed genes present on chromosome 23 in *Leishmania* is shown for *L. major* (Lmaj), *L. donovani* (Ldon), *L. infantum* (Linf), *L. mexicana* (Lmex), *L. braziliensis* (Lbra) and *T. brucei* (Tbru). The loci include several tRNA genes (Ile, Val, Gln, Arg, Leu, Thr, Lys, Gly and Trp), a 5S rRNA gene (5S), two snRNA genes (U1 and U3) and a 7SL RNA gene (7SL). All the species of *Leishmania* show exactly the same order. *T. brucei* shows some degree of synteny. Orthologous genes are joined by grey lines (with the exception of Gly® and Arg* from *T. brucei*, which were not joined to simplify the figure). Flanking protein-coding genes are shown as diagonally striped boxes (not to scale).

Many tRNA genes have not been annotated in the genome databases of *L. infantum*, *L. braziliensis*, *L. mexicana* and *L. donovani*. Thus, in order to compare to *L. major*, we performed *in silico* searches of tRNA genes and other Pol III-transcribed genes. Our results showed that the vast majority of tRNA clusters are highly syntenic among *Leishmania* species. For example, the order of the Pol III-transcribed genes located on chromosome 23 is identical in all the species of *Leishmania* (figure 4). However, several differences were observed. For instance, a cluster of two tRNA genes and a 5S rRNA gene located on chromosome 15 seems to be absent in the corresponding syntenic region of *L. infantum* and *L. donovani*. Also, while cluster I on chromosome 11 [71] contains five tRNA genes and two 5S rRNA genes in *L. major* and *L. mexicana*, it contains only three tRNA genes in *L. infantum*,

For the exclusive use of R. Manning-Cela

L. donovani and *L. braziliensis*. Moreover, several other clusters lack at least one tRNA gene in *L. braziliensis*.

In most eukaryotic organisms, tRNA genes seem to be dispersed randomly throughout the genome. However, in human cells the distribution is non-random, since more than 25% of the tRNA genes are located in a region of only about 4 Mb on chromosome 6 that represents only 0.1% of the human genome [29]. The distribution of genes in the *L. major* genome does not seem to be totally random, since half of the chromosomes do not contain even a single tRNA gene. Additionally, 60 tRNA genes (72%) are located on only seven chromosomes (9, 11, 23, 24, 31, 34 and 36), which represent only 26% of the genome [71].

It was previously shown that transcription of two convergent PGCs on *L. major* chromosome 3 terminates on the convergent strand-switch region, within a tRNA-gene region [59]. Interestingly, 14 of the 39 convergent strand-switch regions (35.9%) in the *L. major* genome contain at least one tRNA gene, representing 45.2% of the 31 tRNA loci. A similar situation was found in *T. brucei*, where 34.6% of the tRNA loci are located within convergent strand-switch regions. This suggests that the use of tRNA genes as signals for termination of transcription of convergent clusters of protein-coding genes might be a common process in trypanosomatids.

The eleven 5S rRNA genes found in the *L. major* genome are distributed on six chromosomes, and are always associated to tRNA genes [20]. A similar organization is present in other species of *Leishmania*. In contrast, the 5S rRNA genes in *T. brucei* and *T. cruzi* are organized into tandem arrays that are not associated to tRNA genes [74; 75]. In trypanosomatids, as in other organisms, the coding regions of the 18S, 5.8S and 28S rRNAs occur as tandem repeats that are clustered at one or several loci. However, a distinctive property of rRNA genes in trypanosomatids is the fragmentation of the 28S-like rRNA into multiple independent molecules: 24Sα, 24Sβ, S1, S2, S4 and S6 [76; 77]. Interestingly, in *L. major* there are two copies of the S4 gene (also known as LSUε) in the majority of the rRNA repeats [78]. In the different species of *Leishmania*, rRNA genes are located on chromosome 27. The number of rRNA repeats varies among species and strains, but it seems to be low (~10), comparing to other eukaryotes (~150) [25; 78; 79].

9. Mitochondrial Genome

As mentioned above, *Leishmania* and other trypanosomatids have only one mitochondrion, and it contains a single kDNA network that is condensed into a disk-shaped structure, positioned in a specialized region of the mitochondrial matrix near the flagellar basal body. The kDNA network is composed of two classes of catenated circular DNA molecules, maxicircles and minicircles [80]. Each mitochondrion has approximately 50 copies of maxicircles, with a size between 20-40 kb, depending on the species. The number of minicircles (~2 kb) ranges from 5,000 to 10,000 per organelle [81]. Maxicircles encode 18 protein-coding genes (components of the respiratory chain), two ribosomal RNAs, and some guide RNAs (gRNAs), which are small RNA molecules that participate in RNA editing. These genes are located on both strands of a ~17 kb-long conservative region [82]. The rest of the maxicircle, known as the divergent region, is composed of repeated sequences.

For the exclusive use of R. Manning-Cela

The minicircle molecules encode from one to five gRNAs. They contain one to four conserved regions and an equal number of variable regions, depending on the species. In *L. tarentolae*, minicircles are organized into a ~170 bp conserved region that contains the origins of replication for both strands, and a variable region that defines the specific minicircle sequence class [83]. The CSB-3 sequence (GGGGTTGGTGTA) (conserved in minicircles from all trypanosomatids) provides a relative position and polarity marker for gRNA genes. A region of bent DNA is situated immediately adjacent to the conserved region. All gRNA genes so far identified are localized within the variable region.

Maxicircle sequences have been reported for *L. tarentolae* and *L. donovani* 1S LdBob. Sequence comparisons showed that the gene order (with the 12S rRNA gene on one side and the ND5 gene on the other) is identical between both species of *Leishmania*, and identical to that of other trypanosomatids [82]. As expected, gene sequences are very similar between *L. tarentolae* and *L. donovani*, with nucleotide identity levels close to 90% [84]. Sequence conservation also included small pre-edited regions of 5'-edited and internally-edited mitochondrial genes, as well as some regions of extensively-edited genes. However, one important difference between both maxicircles is a full-length minicircle insertion found in the 3' region of the ND1 gene of *L. donovani* 1S LdBob [84]. The insertion is 99.1% identical to one of the minicircles from *L. infantum* [85]. The insert has not been observed in other strains and species of *Leishmania*, which suggests that it represents a distinctive characteristic of the strain 1S LdBob of *L. donovani*.

The mitochondrial genome of *Leishmania* is transcribed by a nuclear-encoded mitochondrial RNA polymerase belonging to a family of single-subunit RNA polymerases, as it occurs in other eukaryotes [86]. Neither mitochondrial promoters nor transcription factors have been identified in trypanosomatids [87; 88]. In maxicircles, transcription of the top strand starts ~1.2 kb upstream of the 12S rRNA gene [89], while initiation sites have not been located on the bottom strand. Numerous polycistronic transcripts have been detected, indicating that maxicircles are transcribed polycistronically. Similarly, polycistronic transcripts from minicircles have also been observed [90]. Like maxicircles, no promoters have been identified for minicircles. However, each gRNA transcription unit is flanked by 18-bp imperfect inverted repeats that have been proposed to function in gRNA expression [88].

CONCLUSION

New technologies have allowed the sequencing of the whole genomes of five different species of *Leishmania*. Analyses of the sequences showed that the genomes of these parasites are organized into large DGCs. Comparisons of the genomic sequences revealed a high degree of synteny, not only for protein-coding genes, but also for tRNA genes. Polycistronic transcription of DGCs generates long primary transcripts that are processed by *trans*-splicing and polyadenylation to produce mature mRNAs. Although an unexpectedly small number of species-specific genes was found, gene and chromosome copy number differences were observed between species. Thus, in the absence of regulated transcriptional activity, *Leishmania* seems to raise mRNA abundance by increasing gene-copy number, which can be achieved via gene duplications on disomic chromosomes, or through the formation of supernumerary chromosomes. Changes in gene expression might influence disease tropism.

For the exclusive use of R. Manning-Cela

The *Leishmania* genome contains a small number of tRNA and rRNA genes, and do not contain active retrotransposons (with the exception of *L. braziliensis*). Interestingly, a functional RNAi pathway was identified only in *L. braziliensis* and other members of the *Viannia* subgenus. The potential to manipulate gene expression by RNAi might be very useful for gene function studies in *Leishmania*. Thus, whole-genome sequencing has offered new data that is reshaping research initiatives for leishmaniasis. Further analyses will provide the basis for more detailed molecular studies that could help to improve drug and vaccine design.

Acknowledgments

This work was supported by grants 128461 from CONACyT and IN210712 from PAPIIT (UNAM) to S. Martínez-Calvillo, and by grants 132312 and 139898 from CONACYT, and "proyectos multidisciplinarios" from CINVESTAV-IPN to R. Manning-Cela. N.E. Padilla-Mejía was recipient of a doctoral fellowship from CONACyT.

References

[1] Herwaldt BL. Leishmaniasis. *Lancet,* 1999 354,1191-1199.

[2] Davies, CR; Kaye, P; Croft, SL; Sundar, S. Leishmaniasis: New Approaches To Disease Control. *British Medical Journal*, 2003 326, 377-382.

[3] Antinori, S; Schifanella, L; Corbellino, M. Leishmaniasis: New Insights From An Old And Neglected Disease. *European Journal Of Clinical Microbiology And Infection Diseases,* 2011 Doi 10.1007/S10096-011-1276-0.

[4] Hailu, A; Musa, AM; Royce, C; Wasunna, M. Visceral Leishmaniasis: New Health Tools Are Needed. *Public Library Of Science Medicine*, 2005 2, E211.

[5] Handman E. Cell Biology Of *Leishmania*. *Advances In Parasitology,* 1999 44, 1-39.

[6] WHO. Control Of The Leishmaniases, Technical Report Series 949. Geneva: World Health Organization; 2010, P. 1-186.

[7] Row R. *Leishmania donovani* And *Leishmania tropica*. *British Medical Journal,* 1912 1, 717-718.

[8] Thomson, JG; Sinton, JA. *Leishmania donovani*, In Cultures, Recovered From Spleen During Life And From The Bone Marrow Of The Cadaver. *Proceedings Of The Royal Society Of Medicine,* 1921 14, 21-23.

[9] Banuls, AL; Hide, M; Tibayrenc, M. Evolutionary Genetics And Molecular Diagnosis Of *Leishmania* Species. *Transactions Of The Royal Society Of Tropical Medicine And Hygiene,* 2002 96 Suppl 1, S9-13.

[10] Bates PA. Transmission Of *Leishmania* Metacyclic Promastigotes By Phlebotomine Sand Flies. *International Journal For Parasitology,* 2007 37, 1097-1106.

[11] Schonian, G; Mauricio, I; Cupolillo, E. Is It Time To Revise The Nomenclature Of *Leishmania*? *Trends In Parasitology,* 2010 26, 466-469.

[12] Matthews, KR. The Developmental Cell Biology Of *Trypanosoma brucei*. *Journal Of Cell Science,* 2005 118, 283-290.

For the exclusive use of R. Manning-Cela

[13] Haile, S; Papadopoulou, B. Developmental Regulation Of Gene Expression In Trypanosomatid Parasitic Protozoa. *Current Opinion In Microbiology,* 2007 10; 569-577.

[14] Simpson, L; Thiemann, OH; Savill, NJ; Alfonzo, JD; Maslov, DA. Evolution Of RNA Editing In Trypanosome Mitochondria. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 2000 97; 6986-6993.

[15] Schnaufer, A; Ernst, NL; Palazzo, SS; O'Rear, J; Salavati, R; Stuart, K. Separate Insertion And Deletion Subcomplexes Of The *Trypanosoma brucei* RNA Editing Complex. *Molecular Cell,* 2003 12, 307-319.

[16] Parsons M. Glycosomes: Parasites And The Divergence Of Peroxisomal Purpose. *Molecular Microbiology,* 2004 53, 717-724.

[17] Gull K. The Cytoskeleton Of Trypanosomatid Parasites. *Annual Review Of Microbiology,* 1999 53, 629-655.

[18] Field, MC; Carrington M. The Trypanosome Flagellar Pocket. *Nature Reviews. Microbiology,* 2009 7, 775-786.

[19] Myler, PJ; Audleman, L; Devos, T; Hixson, G; Kiser, P; Lemley, C; et al. *Leishmania major* Friedlin Chromosome 1 Has An Unusual Distribution Of Protein-Coding Genes. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 1999 96, 2902-2906.

[20] Ivens, AC; Peacock, CS; Worthey, EA; Murphy, L; Aggarwal, G; Berriman, M; et al. The Genome Of The Kinetoplastid Parasite, *Leishmania major*. *Science,* 2005 309, 436-442.

[21] El-Sayed, NM; Myler, PJ; Bartholomeu, DC; Nilsson, D; Aggarwal, G; Tran, AN; et al. The Genome Sequence Of *Trypanosoma cruzi*, Etiologic Agent Of Chagas Disease. *Science*, 2005 309, 409-415.

[22] Berriman, M; Ghedin, E; Hertz-Fowler, C; Blandin, G; Renauld, H; Bartholomeu, DC; et al. The Genome Of The African Trypanosome *Trypanosoma brucei*. *Science,* 2005 309, 416-422.

[23] Peacock, CS; Seeger, K; Harris, D; Murphy, L; Ruiz, JC; Quail, MA; et al. Comparative Genomic Analysis Of Three *Leishmania* Species That Cause Diverse Human Disease. *Nature Genetics,* 2007 39, 839-847.

[24] Rogers, MB; Hilley, JD; Dickens, NJ; Wilkes, J; Bates, PA; Depledge, DP; et al. Chromosome And Gene Copy Number Variation Allow Major Structural Change Between Species And Strains Of *Leishmania*. *Genome Research,* 2011 21, 2129-2142.

[25] Downing, T; Imamura, H; Decuypere, S; Clark, TG; Coombs, GH; Cotton, JA; et al. Whole Genome Sequencing Of Multiple *Leishmania donovani* Clinical Isolates Provides Insights Into Population Structure And Mechanisms Of Drug Resistance. *Genome Research,* 2011 21, 2143-2156.

[26] Lukes, J; Mauricio, IL; Schonian, G; Dujardin, JC; Soteriadou, K; Dedet, JP; et al. Evolutionary And Geographical History Of The *Leishmania donovani* Complex With A Revision Of Current Taxonomy. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 2007 104, 9375-9380.

[27] Weatherly, DB; Boehlke, C; Tarleton, RL. Chromosome Level Assembly Of The Hybrid *Trypanosoma cruzi* Genome. *Bmc Genomics,* 2009 10, 255.

For the exclusive use of R. Manning-Cela

[28] Myler, PJ; Beverley, SM; Cruz, AK; Dobson, DE; Ivens, AC; McDonagh, PD; et al. The *Leishmania* Genome Project: New Insights Into Gene Organization And Function. *Medical Microbiology And Immunology*, 2001 190, 9-12.

[29] Lander, ES; Linton, LM; Birren, B; Nusbaum, C; Zody, MC; Baldwin, J; et al. Initial Sequencing And Analysis Of The Human Genome. *Nature*, 2001 409, 860-921.

[30] Mair, G; Shi, H; Li, H; Djikeng, A; Aviles, HO; Bishop, JR; et al. A New Twist In Trypanosome RNA Metabolism: Cis-Splicing Of Pre-mRNA. *RNA*, 2000 6, 163-169.

[31] Bringaud, F; Ghedin, E; El-Sayed, NM; Papadopoulou, B. Role Of Transposable Elements In Trypanosomatids. *Microbes And Infecttion*, 2008 10, 575-581.

[32] Ghedin, E; Bringaud, F; Peterson, J; Myler, P; Berriman, M; Ivens, A; et al. Gene Synteny And Evolution Of Genome Architecture In Trypanosomatids. *Molecular And Biochemical Parasitology*, 2004 134, 183-191.

[33] Bringaud, F; Muller, M; Cerqueira, GC; Smith, M; Rochette, A; El-Sayed, NM; et al. Members Of A Large Retroposon Family Are Determinants Of Post-Transcriptional Gene Expression In *Leishmania*. *Public Library Of Science Pathogens*, 2007 3, 1291-1307.

[34] Requena, JM; Folgueira, C; Lopez, MC; Thomas, MC. The Sider2 Elements, Interspersed Repeated Sequences That Populate The *Leishmania* Genomes, Constitute Subfamilies Showing Chromosomal Proximity Relationship. *Bmc Genomics*, 2008 9, 263.

[35] Teixeira, SM; De Paiva, RM; Kangussu-Marcolino, MM; Darocha, WD. Trypanosomatid Comparative Genomics: Contributions To The Study Of Parasite Biology And Different Parasitic Diseases. *Genetics And Molecular Biology*, 2012 35, 1-17.

[36] Smith, DF; Peacock, CS; Cruz, AK. Comparative Genomics: From Genotype To Disease Phenotype In The Leishmaniases. *International Journal For Parasitology*, 2007 37, 1173-1186.

[37] Zhang, WW; Mendez, S; Ghosh, A; Myler, P; Ivens, A; Clos, J; et al. Comparison Of The A2 Gene Locus In *Leishmania donovani* And *Leishmania major* And Its Control Over Cutaneous Infection. *Journal Of Biological Chemistry*, 2003 278, 35508-35515.

[38] Fire, A; Xu, S; Montgomery, MK; Kostas, SA; Driver, SE; Mello, CC. Potent And Specific Genetic Interference By Double-Stranded RNA In *Caenorhabditis elegans*. *Nature*, 1998 391, 806-811.

[39] Ngo, H; Tschudi, C; Gull, K; Ullu, E. Double-Stranded RNA Induces mRNA Degradation In *Trypanosoma brucei*. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 1998 95, 14687-14692.

[40] Robinson, KA; Beverley, SM. Improvements In Transfection Efficiency And Tests Of RNA Interference (RNAi) Approaches In The Protozoan Parasite *Leishmania*. *Molecular And Biochemical Parasitology*, 2003 128, 217-228.

[41] Lye, LF; Owens, K; Shi, H; Murta, SM; Vieira, AC; Turco, SJ; et al. Retention And Loss Of RNA Interference Pathways In Trypanosomatid Protozoans. *Public Library Of Science Pathogens*, 2010 6, E1001161.

[42] Britto, C; Ravel, C; Bastien, P; Blaineau, C; Pages, M; Dedet, JP; et al. Conserved Linkage Groups Associated With Large-Scale Chromosomal Rearrangements Between Old World And New World *Leishmania* Genomes. *Gene*, 1998 222, 107-117.

For the exclusive use of R. Manning-Cela

[43] Bastien, P; Blaineau, C; Pages, M. *Leishmania*: Sex, Lies And Karyotype. *Parasitology Today,* 1992, 8; 174-177.

[44] Martinez-Calvillo, S; Stuart, K; Myler, PJ. Ploidy Changes Associated With Disruption Of Two Adjacent Genes On *Leishmania major* Chromosome 1. *International Journal For Parasitology,* 2005 35, 419-429.

[45] Sunkin, SM; Kiser, P; Myler, PJ; Stuart, K. The Size Difference Between *Leishmania major* Friedlin Chromosome One Homologues Is Localized To Sub-Telomeric Repeats At One Chromosomal End. *Molecular And Biochemical Parasitology,* 2000 109, 1-15.

[46] Sterkers, Y; Lachaud, L; Crobu, L; Bastien, P; Pages, M. Fish Analysis Reveals Aneuploidy And Continual Generation Of Chromosomal Mosaicism In *Leishmania major*. *Cellular Microbiology,* 2011 13, 274-283.

[47] Rochette, A; McNicoll, F; Girard, J; Breton, M; Leblanc, E; Bergeron, MG; et al. Characterization And Developmental Gene Regulation Of A Large Gene Family Encoding Amastin Surface Proteins In *Leishmania* Spp. *Molecular And Biochemical Parasitology,* 2005 140, 205-220.

[48] Mauricio, IL; Gaunt, MW; Stothard, JR; Miles, MA. Glycoprotein 63 (Gp63) Genes Show Gene Conversion And Reveal The Evolution Of Old World *Leishmania*. *International Journal For Parasitology,* 2007 37, 565-576.

[49] Devault, A; Banuls, AL. The Promastigote Surface Antigen Gene Family Of The *Leishmania* Parasite: Differential Evolution By Positive Selection And Recombination. *Bmc Evolutionary Biology,* 2008 8, 292.

[50] Johnson, PJ; Kooter, JM; Borst, P. Inactivation Of Transcription By UV Irradiation Of *T. brucei* Provides Evidence For A Multicistronic Transcription Unit Including A VSG Gene. *Cell,* 1987 51, 273-281.

[51] Mottram, JC; Murphy, WJ; Agabian, N. A Transcriptional Analysis Of The *Trypanosoma brucei* Hsp83 Gene Cluster. *Molecular And Biochemical Parasitology,* 1989 37, 115-127.

[52] Martinez-Calvillo, S; Yan, S; Nguyen, D; Fox, M; Stuart, K; Myler, PJ. Transcription Of *Leishmania major* Friedlin Chromosome 1 Initiates In Both Directions Within A Single Region. *Molecular Cell,* 2003 11, 1291-1299.

[53] Campbell, DA; Thomas, S; Sturm, NR. Transcription In Kinetoplastid Protozoa: Why Be Normal? *Microbes And Infection,* 2003 5, 1231-1240.

[54] Martinez-Calvillo, S; Vizuet-de-Rueda, JC; Florencio-Martinez, LE; Manning-Cela, RG; Figueroa-Angulo, EE. Gene Expression In Trypanosomatid Parasites. *Journal Of Biomedicine And Biotechnology,* 2010, 525241.

[55] Parsons, M; Nelson, RG; Watkins, KP; Agabian, N. Trypanosome mRNAs Share A Common 5' Spliced Leader Sequence. *Cell,* 1984 38, 309-316.

[56] Kooter, JM; Borst, P. Alpha-Amanitin-Insensitive Transcription Of Variant Surface Glycoprotein Genes Provides Further Evidence For Discontinuous Transcription In Trypanosomes. *Nucleic Acids Research,* 1984 12, 9457-9472.

[57] Clayton, CE. Life Without Transcriptional Control? From Fly To Man And Back Again. *The Embo Journal,* 2002 21, 1881-1888.

[58] Boucher, N; Wu, Y; Dumas, C; Dube, M; Sereno, D; Breton, M; et al. A Common Mechanism Of Stage-Regulated Gene Expression In *Leishmania* Mediated By A Conserved 3'-Untranslated Region Element. *Journal Of Biological Chemistry,* 2002 277, 19511-19520.

For the exclusive use of R. Manning-Cela

[59] Martinez-Calvillo, S; Nguyen, D; Stuart, K; Myler, PJ. Transcription Initiation And Termination On *Leishmania major* Chromosome 3. *Eukaryotic Cell*, 2004 3, 506-517.

[60] Thomas, S; Green, A; Sturm, NR; Campbell, DA; Myler, PJ. Histone Acetylations Mark Origins Of Polycistronic Transcription In *Leishmania major*. *Bmc Genomics* 2009 10, 152.

[61] Saito, RM; Elgort, MG; Campbell, DA. A Conserved Upstream Element Is Essential For Transcription Of The *Leishmania tarentolae* Mini-Exon Gene. *The Embo Journal*, 1994 13, 5460-5469.

[62] Gunzl, A; Ullu, E; Dorner, M; Fragoso, SP; Hoffmann, KF; Milner, JD; et al. Transcription Of The *Trypanosoma brucei* Spliced Leader RNA Gene Is Dependent Only On The Presence Of Upstream Regulatory Elements. *Molecular And Biochemical Parasitology*, 1997 85, 67-76.

[63] Luo, H; Bellofatto, V. Characterization Of Two Protein Activities That Interact At The Promoter Of The Trypanosomatid Spliced Leader RNA. *Journal Of Biological Chemistry*, 1997 272, 33344-33352.

[64] Gunzl, A; Vanhamme, L; Myler, PJ. Transcription In Trypanosomes: A Different Means To The End. Barry, JD; McCulloch, R; Mottram, JC; Acosta-Serrano, A. (Ed.) In: Trypanosomes: After The Genome. Wymonham: Horizon Bioscience; 2007; P. 177-208.

[65] Hernandez-Rivas, R; Florencio-Martinez, LE; Martinez-Salazar, M; Martinez-Calvillo, S. Gene Expression And Transcriptional Machinery In Trypanosomatid And Apicomplexa Parasites. Terrazas, LI. (Ed.). In: Advances In The Immunobiology Of Parasitic Diseases. Kerala, India: Research Signpost; 2007; P. 313-37.

[66] Ruan, JP; Arhin, GK; Ullu, E; Tschudi, C. Functional Characterization Of A *Trypanosoma brucei* TATA-Binding Protein-Related Factor Points To A Universal Regulator Of Transcription In Trypanosomes. *Molecular And Cellular Biology*, 2004 24, 9610-9618.

[67] Palenchar, JB; Liu, W; Palenchar, PM; Bellofatto, V. A Divergent Transcription Factor TFIIB In Trypanosomes Is Required For RNA Polymerase II-Dependent Spliced Leader RNA Transcription And Cell Viability. *Eukaryotic Cell*, 2006 5, 293-300.

[68] Schimanski, B; Brandenburg, J; Nguyen, TN; Caimano, MJ; Gunzl, A. A TFIIB-Like Protein Is Indispensable For Spliced Leader RNA Gene Transcription In *Trypanosoma brucei*. *Nucleic Acids Research*, 2006 34, 1676-1684.

[69] Lecordier, L; Devaux, S; Uzureau, P; Dierick, JF; Walgraffe, D; Poelvoorde, P; et al. Characterization Of A TFIIH Homologue From *Trypanosoma brucei*. *Molecular Microbiology*, 2007 64, 1164-1181.

[70] Lee, JH; Nguyen, TN; Schimanski, B; Gunzl, A. Spliced Leader RNA Gene Transcription In *Trypanosoma brucei* Requires Transcription Factor TFIIH. *Eukaryotic Cell*, 2007 6, 641-649.

[71] Padilla-Mejia, NE; Florencio-Martinez, LE; Figueroa-Angulo, EE; Manning-Cela, RG; Hernandez-Rivas, R; Myler, PJ et al. Gene Organization And Sequence Analyses Of Transfer RNA Genes In Trypanosomatid Parasites. *Bmc Genomics*, 2009 10, 232.

[72] Marck, C; Grosjean, H. tRNomics: Analysis Of tRNA Genes From 50 Genomes Of Eukarya, Archaea, And Bacteria Reveals Anticodon-Sparing Strategies And Domain-Specific Features. *Rna*, 2002 8, 1189-1232.

For the exclusive use of R. Manning-Cela

[73] Goodenbour, JM; Pan, T. Diversity Of tRNA Genes In Eukaryotes. *Nucleic Acids Research*, 2006 34, 6137-6146.

[74] Hasan, G; Turner, MJ; Cordingley, JS. Ribosomal RNA Genes Of *Trypanosoma brucei*: Mapping The Regions Specifying The Six Small Ribosomal RNAs. *Gene* 1984 27, 75-86.

[75] Hernandez-Rivas, R; Martinez-Calvillo, S; Romero, M; Hernandez R. *Trypanosoma cruzi* 5S rRNA Genes: Molecular Cloning, Structure And Chromosomal Organization. *The Fems Microbiology Letters*, 1992 71, 63-67.

[76] Spencer, DF; Collings, JC; Schnare, MN; Gray, MW. Multiple Spacer Sequences In The Nuclear Large Subunit Ribosomal RNA Gene Of *Crithidia fasciculata*. *The Embo Journal*, 1987 6, 1063-1071.

[77] Hernandez, R; Diaz De Leon, F; Castaneda, M. Molecular Cloning And Partial Characterization Of Ribosomal RNA Genes From *Trypanosoma cruzi*. *Molecular And Biochemical Parasitology*, 1988 27, 275-279.

[78] Martinez-Calvillo, S; Sunkin, SM; Yan, S; Fox, M; Stuart, K; Myler, PJ. Genomic Organization And Functional Characterization Of The *Leishmania major* Friedlin Ribosomal RNA Gene Locus. *Molecular And Biochemical Parasitology*, 2001 116, 147-157.

[79] Inga, R; De Doncker, S; Gomez, J; Lopez, M; Garcia, R; Le Ray, D; et al. Relation Between Variation In Copy Number Of Ribosomal RNA Encoding Genes And Size Of Harbouring Chromosomes In *Leishmania* Of Subgenus Viannia. *Molecular And Biochemical Parasitology*, 1998 92, 219-228.

[80] Liu, B; Liu, Y; Motyka, SA; Agbo, EE; Englund, PT. Fellowship Of The Rings: The Replication Of Kinetoplast DNA. *Trends In Parasitology*, 2005 21, 363-369.

[81] Lukes, J; Guilbride, DL; Votypka, J; Zikova, A; Benne, R; Englund, PT. Kinetoplast DNA Network: Evolution Of An Improbable Structure. *Eukaryotic Cell*, 2002 1, 495-502.

[82] Simpson, L; Wang, SH; Thiemann, OH; Alfonzo, JD; Maslov, DA; Avila, HA. U-Insertion/Deletion Edited Sequence Database. *Nucleic Acids Research*, 1998 26, 170-176.

[83] Simpson, L. The Genomic Organization Of Guide RNA Genes In Kinetoplastid Protozoa: Several Conundrums And Their Solutions. *Molecular And Biochemical Parasitology*, 1997 86, 133-141.

[84] Nebohacova, M; Kim, CE; Simpson, L; Maslov, DA. RNA Editing And Mitochondrial Activity In Promastigotes And Amastigotes Of *Leishmania donovani*. *International Journal For Parasitology*, 2009 39, 635-644.

[85] Tripp, CA; Myler, PJ; Stuart, KD. Nucleotide Sequence Of A Minicircle From *Leishmania infantum*. *Molecular And Biochemical Parasitology*, 1993 62, 319-320.

[86] Clement, SL; Koslowsky, DJ. Unusual Organization Of A Developmentally Regulated Mitochondrial RNA Polymerase (Tbmtrnap) Gene In *Trypanosoma brucei*. *Gene*, 2001 272, 209-218.

[87] Lukes, J; Hashimi, H; Zikova, A. Unexplained Complexity Of The Mitochondrial Genome And Transcriptome In Kinetoplastid Flagellates. *Current Genetics*, 2005 48, 277-299.

[88] Aphasizhev, R; Aphasizheva, I. Mitochondrial RNA Processing In Trypanosomes. *Research In Microbiology*, 2011 162, 655-663.

For the exclusive use of R. Manning-Cela

[89] Michelotti, EF; Harris, ME; Adler, B; Torri, AF; Hajduk, SL. *Trypanosoma brucei* Mitochondrial Ribosomal RNA Synthesis, Processing And Developmentally Regulated Expression. *Molecular And Biochemical Parasitology,* 1992 54, 31-41.
[90] Grams, J; McManus, MT; Hajduk, SL. Processing Of Polycistronic Guide RNAs Is Associated With RNA Editing Complexes In *Trypanosoma brucei*. *The Embo Journal,* 2000 19, 5525-5532.

For the exclusive use of R. Manning-Cela